

ÉCOLE DE TECHNOLOGIE SUPÉRIEURE
UNIVERSITÉ DU QUÉBEC

THÈSE PAR ARTICLES PRÉSENTÉE À
L'ÉCOLE DE TECHNOLOGIE SUPÉRIEURE

COMME EXIGENCE PARTIELLE
À L'OBTENTION DU
DOCTORAT EN GÉNIE
Ph.D.

PAR
KHALILI MALIKA

NOUVELLE APPROCHE DE GÉNÉRATION MULTI-SITE DES DONNÉES
CLIMATIQUES

MONTREAL, LE 5 SEPTEMBRE 2007

© Khalili Malika, 2007

CETTE THÈSE A ÉTÉ ÉVALUÉE

PAR UN JURY COMPOSÉ DE :

M. Robert Leconte, directeur de thèse
Département de génie de la construction à l'École de technologie supérieure

M. François Brissette, codirecteur de thèse
Département de génie de la construction à l'École de technologie supérieure

M. Jean-Sébastien Dubé, président du jury
Département de génie de la construction à l'École de technologie supérieure

M. Jean Rousselle, membre du jury
Département des génies civil, géologique et des mines à l'École polytechnique

M. Taha B.M. J Ouarda, membre du jury
Centre Eau, Terre et Environnement à l'Institut national de la recherche scientifique

ELLE A FAIT L'OBJET D'UNE SOUTENANCE DEVANT JURY ET PUBLIC

LE 4 SEPTEMBRE 2007

À L'ÉCOLE DE TECHNOLOGIE SUPÉRIEURE

REMERCIEMENTS

Je tiens à adresser mes remerciements et ma sincère reconnaissance à mon directeur de recherche, le professeur ROBERT LECONTE, et co-directeur, le professeur FRANÇOIS BRISSETTE, de m'avoir accueillie au sein du laboratoire DRAME (Développement et Recherche Appliquée à la Modélisation Environnementale) à l'École de technologie supérieure. Je les remercie pour leur soutien, leurs encouragements et pour le financement qu'ils m'ont octroyé tout au long de cette thèse ainsi que pour les opportunités qu'ils m'ont accordées pour participer à des conférences nationales et internationales.

Je remercie les membres du Jury qui ont accepté d'évaluer cette thèse.

Je remercie également l'École de technologie supérieure pour les bourses qu'elle m'a accordées. Je remercie tous mes professeurs, d'ici et d'ailleurs, pour les formations qu'ils m'ont apportées.

Mes remerciements vont également au Conseil de Recherche en Sciences Naturelles et en Génie du Canada (CRSNG), Hydro-Québec et le consortium, Ouranos, sur la climatologie régionale et l'adaptation aux changements climatiques, qui ont financé le projet mené au DRAME dont cette thèse fait partie. Je remercie Ouranos de m'avoir donné l'opportunité de présenter mes résultats de recherche aux conférences et symposiums qu'il organise régulièrement.

Je tiens également à remercier les réviseurs anonymes du Journal of Hydrometeorology et Journal of Hydrology pour leurs commentaires et critiques qui ont permis d'améliorer les versions initiales de mes articles.

Je remercie mes collègues du laboratoire DRAME pour le climat favorable et respectueux qu'ils y ont toujours établi.

La présente thèse est avant tout dédiée à mes parents, FATIMA et MOHAMED, à qui j'adresse ma plus profonde reconnaissance et affection pour leur éducation, leurs prières, leur soutien, leurs encouragements, leurs conseils et leurs sacrifices qui m'ont permis d'arriver là.

À mon mari, ABDOU, qui a enduré les exigences et les contraintes de cette thèse, je tiens à lui adresser mes plus vifs remerciements pour sa patience, son soutien et ses encouragements infinis. Merci ABDOU, cette thèse est aussi la tienne.

Finalement, j'adresse mes forts remerciements à toute ma famille, notamment, ma sœur AICHA qui a tant veillé à ce que je mène à bien mes études et mon beau frère MOHAMED OUGABRAI qui a su me soutenir, susciter la confiance en moi et m'encourager durant tout mon cursus scolaire. Qu'ils trouvent ici l'expression de ma grande reconnaissance.

NOUVELLE APPROCHE DE GÉNÉRATION MULTI-SITE DES DONNÉES CLIMATIQUES

KHALILI, Malika

RÉSUMÉ

Les générateurs de climat peuvent être utilisés pour des études hydrologiques, agraires, environnementales et pour évaluer les impacts potentiels des changements climatiques. Ces outils statistiques permettent de générer de longues séries temporelles de données climatiques statistiquement similaires aux données observées. Afin de générer le climat futur, leurs paramètres peuvent être modifiés pour tenir compte des scénarios de changements climatiques issus des modèles de circulation générale (MCG). À cause de la faible résolution spatiale des MCG, les générateurs de climat constituent des techniques statistiques prometteuses de mise à l'échelle pour développer des scénarios climatiques locaux.

Malgré l'abondance des générateurs de climat, un nombre très limité a été conçu pour tenir compte de la dépendance spatiale qui caractérise réellement les données climatiques. Toutefois, la complexité des générateurs de climat multi-sites existants et le nombre élevé de paramètres qu'ils font intervenir ont conduit à une réticence envers ce type de générateurs au profit des générateurs uni-site, plus faciles à implanter et à utiliser. Or, la dépendance spatiale ne peut nullement être négligée pour des études où elle a un effet primordial, comme en hydrologie.

Le présent travail doctoral a permis de développer une approche de régionalisation, pratique et efficace, basée sur le concept d'autocorrélation spatiale. Cette dernière permet de décrire la dépendance spatiale des valeurs d'une variable distribuée spatialement. Ce concept est utilisé dans cette thèse de façon à reproduire simultanément, dans les séries synthétiques de données climatiques, les autocorrélations spatiales observées sur l'ensemble des stations météorologiques et les corrélations observées entre chaque couple de stations.

L'approche de génération multi-site est appliquée avec succès sur le bassin versant de la rivière Péribonca situé sur la province du Québec, Canada. Elle est initialement appliquée pour la génération multi-site des données journalières d'occurrences et de quantités de précipitations. Une extension est ensuite réalisée pour les données journalières de températures maximales, minimales et de radiations solaires. Une étude de modélisation hydrologique comparative est effectuée en utilisant le générateur multi-site développé et le générateur uni-site de base. Le modèle hydrologique distribué, Hydrotel, est sélectionné pour évaluer l'effet de la dépendance spatiale des données climatiques sur l'efficacité de la modélisation hydrologique. Les résultats confirment la supériorité du générateur de climat multi-site dans la simulation des débits moyens et des fréquences et magnitudes des crues extrêmes estivales, automnales et printanières.

NEW MULTI-SITE GENERATION APPROACH OF CLIMATE DATA

KHALILI, Malika

ABSTRACT

Weather generators can be used for various water resources applications in hydrology, agriculture, environment and climate change studies. They are being increasingly used to address these issues. These statistical tools allow the generation of long time series of climate data with the same statistical properties as the observed ones. Weather generator parameters can be modified, using climate change projections from global circulation models (GCM), to provide future climate time series. Due to the low spatial resolution of GCMs, weather generators can also be used as statistical downscaling tools to create local scale climate scenarios.

Most of the existing weather generators operate at a single site independently of the others and thus ignore the spatial dependence exhibited by the observed data. The few multi-site weather generators are complex and involve large number of parameters, which makes the uni-site weather generators preferred over multi-site generators because of their easier implementation and use. However, the spatial dependence between the weather stations should be preserved in studies where it has a crucial effect, such as hydrology.

The present doctoral work developed an efficient regionalisation approach, easy to implement, based on the spatial autocorrelation concept. The latter describes the dependence among values of a variable taking into account their spatial location. This thesis uses spatial autocorrelation to reproduce, in the synthetic climate time series, both the spatial autocorrelations observed over the set of weather stations and the correlations between each pair of stations.

The multi-site generation approach is successfully applied to the Péribonca River Basin in the Canadian province of Quebec. It is firstly implemented for the multi-site generation of daily occurrence and amount precipitation data and then extended to the multi-site generation of daily maximum temperature, minimum temperature, and solar radiation data. A comparative hydrological modelling study is carried out using the developed multi-site weather generator and the basic uni-site weather generator. The distributed hydrological model, Hydrotel, is selected to assess the effect of climate data spatial dependence on the effectiveness of the hydrological modelling. Results demonstrate the efficiency of the multi-site weather generator in the simulation of average streamflows as well as the frequencies and magnitudes of summer, autumn and spring floods.

TABLE DES MATIÈRES

	Page
INTRODUCTION	1
CHAPITRE 1 REVUE BIBLIOGRAPHIQUE	5
1.1 Introduction.....	5
1.2 Revue bibliographique sur les générateurs de climat	6
1.2.1 Générateurs de climat paramétriques uni-site.....	6
1.2.2 Générateurs de climat paramétriques multi-sites.....	9
1.2.3 Générateurs de climat non paramétriques uni-site.....	12
1.2.4 Générateurs de climat non paramétriques multi-sites.....	15
1.3 Conclusion	16
CHAPITRE 2 DESCRIPTION GÉNÉRALE DES PROCESSUS DE GÉNÉRATION UNI-SITE ET DE L'APPROCHE DE RÉGIONALISATION	18
2.1 Description des processus de génération uni-site	18
2.1.1 Génération des occurrences de précipitations.....	18
2.1.2 Génération des quantités de précipitations	21
2.1.3 Génération des autres variables climatiques	23
2.2 Approche de régionalisation : l'autocorrélation spatiale	24
2.2.1 Qu'est ce que l'autocorrélation spatiale ?.....	24
2.2.2 Pourquoi l'autocorrélation spatiale ?	27
CHAPITRE 3 ARTICLE 1: STOCHASTIC MULTI-SITE GENERATION OF DAILY PRECIPITATION DATA USING SPATIAL AUTOCORRELATION.....	31
3.1 Introduction.....	33
3.2 Spatial autocorrelation definition.....	35
3.3 Methodology	38
3.3.1 Uni-site weather generator.....	38
3.3.2 Proposed multi-site generation approach.....	39
3.3.2.1 Spatially autocorrelated random numbers model	39
3.3.2.2 Moving average coefficients estimation	41
3.3.2.3 Data generation	43
3.4 Description of data set	43
3.5 Results and discussion	46
3.6 Conclusion	51
CHAPITRE 4 ARTICLE 2: STOCHASTIC MULTI-SITE GENERATION OF DAILY WEATHER DATA.....	76
4.1 Introduction.....	77
4.2 Multi-site generation approach of daily precipitation data.....	80
4.2.1 Definition of spatial autocorrelation	80
4.2.2 Multi-site generation with spatial intermittence approach.....	82

4.3	Uni-site generation approach of daily temperature and solar radiation data	84
4.4	Multi-site generation approach of daily temperature and solar radiation data	86
4.4.1	Spatially autocorrelated random numbers model	86
4.4.2	Multi-site temperature and solar radiation model	87
4.4.3	Moving average coefficients estimation	88
4.5	Results and discussion	90
4.5.1	Results of spatial intermittence approach	90
4.5.2	Results of multi-site generation approach of daily temperature data	92
4.6	Conclusion	95
CHAPITRE 5 ARTICLE 3: EFFECTIVENESS OF MULTI-SITE WEATHER GENERATOR FOR HYDROLOGICAL MODELLING: COMPARATIVE STUDY WITH UNI-SITE WEATHER GENERATOR USING DISTRIBUTED HYDROLOGICAL MODEL		
5.1	Introduction.....	114
5.2	Uni-site weather generator	117
5.2.1	Uni-site generation of daily precipitation data.....	117
5.2.2	Uni-site generation of daily temperature and solar radiation data	118
5.3	Multi-site weather generator (Khalili <i>et al.</i> 2007, 2008a)	119
5.3.1	Multi-site generation of daily precipitation data.....	119
5.3.2	Multi-site generation of daily temperature and solar radiation data	121
5.4	Hydrological modeling	122
5.4.1	Description of the study area	122
5.4.2	Hydrotel model application.....	123
5.5	Results and discussion	124
5.6	Conclusion	129
CONCLUSION		146
RECOMMANDATIONS		152
BIBLIOGRAPHIE		154

LISTE DES TABLEAUX

	Page
Tableau 3.1 Location and recorded years of the used stations	73
Tableau 3.2 Observed numbers of wet days per month per station.....	74
Tableau 3.3 Monthly average amounts of precipitation observed at each station.....	75
Tableau 4.1 Location and recorded years of the used nine stations	111
Tableau 5.1 Soil type on Chute du Diable Watershed	141
Tableau 5.2 Land-use occupation on Chute du Diable Watershed.....	142
Tableau 5.3 Summer flood frequency using observed data, multi-site, and uni-site generated data in Hydrotel model	143
Tableau 5.4 Autumn flood frequency using observed data with outliers (30 pieces of data) and without outliers (28 pieces of data), multi-site, and uni-site generated data in Hydrotel model	144
Tableau 5.5 Spring flood frequency using observed data, multi-site, and uni-site generated data in Hydrotel model	145

LISTE DES FIGURES

	Page
Figure 2.1	Types d'autocorrélations spatiales.....26
Figure 3.1	Flow chart for the proposed multi-site generation approach54
Figure 3.2	The Peribonca watershed, showing the location of meteorological stations. Elevations in meters.....55
Figure 3.3	Relationship between the moving average coefficients γ and Moran's I_X computed for the precipitation occurrences at the seven stations for June56
Figure 3.4	Observed versus simulated monthly numbers of rainy days at all stations for January, April, July and October.57
Figure 3.5	Joint probabilities from observed and simulated occurrence process for each pair of stations and each month. Figure 5a for both stations are dry and figure 5b for both stations are wet.58
Figure 3.6	Observed and simulated daily spatial autocorrelations I_X of precipitation occurrences.59
Figure 3.7	Relationship between the moving average coefficients γ and Moran's I_R computed for the precipitation amounts at the seven stations for January60
Figure 3.8	Observed and simulated total monthly precipitations at the seven stations for February, June, September and December.....61
Figure 3.9	Observed and simulated daily spatial autocorrelations I_R of precipitation amounts.....62
Figure 3.10	Observed and simulated daily precipitation standard deviations at the seven stations for January, April, August and October.63

Figure 3.11	Observed and simulated interstation correlations between daily precipitation occurrences for all pairs of stations and months.....	64
Figure 3.12	Observed and simulated interstation correlations between daily precipitation amounts for all pairs of stations and months.	65
Figure 3.13	Observed and simulated interstation correlations between daily precipitation occurrences for all pairs of stations and months using monthly weight matrices.....	66
Figure 3.14	Observed and simulated interstation correlations between daily precipitation amounts for all pairs of stations and months using monthly weight matrices.....	67
Figure 3.15	Relationship between precipitation occurrence correlations and random number correlations for Péribonca and Chute du Diable stations in January.....	68
Figure 3.16	Observed and simulated interstation correlations between daily precipitation occurrences for all pairs of stations and months using the Wilks' approach	69
Figure 3.17	Observed and simulated interstation correlations between daily precipitation amounts for all pairs of stations and months using the Wilks' approach.	70
Figure 3.18	Observed and simulated monthly spatial autocorrelations of precipitation occurrences using (a) the proposed approach and (b) the Wilks' approach.....	71
Figure 3.19	Observed and simulated monthly spatial autocorrelations of precipitation amounts using (a) the proposed approach and (b) the Wilks' approach.	72
Figure 4.1	Flow chart for the proposed multi-site generation approach of daily temperature and solar radiation data.....	97
Figure 4.2	The Peribonca watershed, showing the location of meteorological stations. Elevations in meters.....	98

Figure 4.3	Relationship between the mean precipitation amounts (mm) of Peribonca station and the spatial dependence values of occurrences over the watershed in September99
Figure 4.4	Spatial intermittence using the simple exponential function (a) and the spatial exponential functions (b)100
Figure 4.5	Relationship between the moving average coefficients $\gamma_{T_{\max}}$ and the spatial autocorrelations $I_{T_{\max}}$ for maximum temperature data in July101
Figure 4.6	Observed and simulated daily spatial autocorrelations of maximum temperature (a) and minimum temperature (b).....102
Figure 4.7	Observed versus simulated interstation correlations of daily maximum temperature (a) and minimum temperature (b) between each pair of the seven stations and month.103
Figure 4.8	Observed versus simulated interstation correlations of daily maximum temperature (a) and minimum temperature (b) between each pair of the seven stations and month using the monthly weight matrices104
Figure 4.9	Observed and simulated daily spatial autocorrelations of maximum temperature (a) and minimum temperature (b) using the monthly weight matrices.105
Figure 4.10	Observed versus simulated monthly means of daily maximum temperature (a) and minimum temperature (b).....106
Figure 4.11	Observed versus simulated interstation correlations of daily maximum temperature (a) and minimum temperature (b) between each pair of the nine stations and month using the monthly weight matrices.107
Figure 4.12	Observed and simulated daily spatial autocorrelations of maximum temperature (a) and minimum temperature (b) using the nine stations and the monthly weight matrices.108
Figure 4.13	Observed versus simulated interstation correlations of daily maximum temperature (a) and minimum temperature (b) between each pair of the nine stations and month using the monthly weight matrices and <i>SDI</i>109

Figure 4.14	Observed and simulated daily spatial dependence of maximum temperature (a) and minimum temperature (b) using the monthly weight matrices and <i>SDI</i>110
Figure 5.1	Chute du Diable watershed with locations of weather stations131
Figure 5.2	River reaches and altitudes (m) of the Chute du Diable watershed obtained from Physitel132
Figure 5.3	Reconstituted and simulated streamflows using multi-site and uni-site weather generators133
Figure 5.4	Spatial distribution of heavy total rainfall amounts occurring in September over the Chute du Diable watershed from (a) observed data, (b) multi-site generated data, and (c) uni-site generated data134
Figure 5.5	Summer streamflow frequency distributions using (a) observed data, (b) multi-site generated data, and (c) uni-site generated data135
Figure 5.6	Autumn streamflow frequency distributions using (a) observed data, (b) multi-site generated data, and (c) uni-site generated data136
Figure 5.7	Spring streamflow frequency distributions using (a) observed data, (b) multi-site generated data, and (c) uni-site generated data137
Figure 5.8	Monthly streamflow averages.....138
Figure 5.9	Monthly streamflow standard deviations.....139
Figure 5.10	Monthly air temperature standard deviations140

LISTE DES ABRÉVIATIONS, SIGLES ET ACRONYMES

AIC	Critère d'information d'Akaike
ALCAN	Compagnie d'aluminium du Canada
AR1	Modèle autorégressif d'ordre 1
B	Station météorologique Bonnard
BIC	Critère d'information bayésien
BV3C	Bilan vertical en trois couches
Cd	Station météorologique Chute-du-Diable
Cp	Station météorologique Chute-des-Passes
CRSNG	Conseil de Recherche en Sciences Naturelles et en Génie du Canada
DEM	Modèle d'élévation numérique
DRAME	Développement et Recherche Appliquée à la Modélisation Environnementale
Eq.	Équation
GIEC	Groupe d'experts intergouvernemental sur l'évolution du climat
H	Station météorologique Hémon
LARS-WG	Générateur de climat développé par le « Long-Ashton Research Station »
MCG	Modèle de Circulation Générale
N	Station météorologique Normandin CDA
NHMM	Modèles de Markov cachés non homogènes
P	Station météorologique Péribonca
RHHU	Unités hydrologiques relativement homogènes
SI	Station météorologique St-Léon-de-Labrecque
SWAT	Outil d'évaluation de sol et d'eau

USA	États-Unis d'Amérique
WeaGETS	Générateur de climat développé à l'École de Technologie Supérieure
WGEN	Générateur de climat développé par Richardson (1981)
WMO	Organisation météorologique mondiale

LISTE DES SYMBOLES

A	Matrice (3×3) fonction des matrices M_0 et M_1 .
B	Matrice (3×3) fonction des matrices M_0 et M_1 .
C	Indice de Geary
D_t	Vecteur d'état utilisé par les générateurs de climat non paramétriques
D_i	Vecteur d'état actuel
$D_{j(i)}$	$j^{ème}$ proche voisin de D_i
E	Espérance statistique
F	Fonction de densité cumulative
I	Indice de Moran
$I_{T\max}$	Indice de Moran calculé pour les températures maximales
$I_{T\min}$	Indice de Moran calculé pour les températures minimales
I_{Sr}	Indice de Moran calculé pour les radiations solaires
$J_{i,k}$	Ensemble ordonné des k proches voisins de D_i
$k[j(i)]$	Kernel discret
l	Nombre total de jours d'un mois
m	Nombre total des valeurs de γ prises de son intervalle de variation
M_0	Matrice d'autocorrélations et correlations croisées de lag 0 des résidus observés de températures maximales, températures minimales et de radiations solaires
M_1	Matrice d'autocorrélations et correlations croisées de lag 1 des résidus observés de températures maximales, températures minimales et de radiations solaires
n	Nombre de sites
n_{01}	Nombre des 0 suivis par 1
n_{11}	Nombre des 1 suivis par 1
$n_{0.}$	Nombre des 0 suivis par une autre valeur
$n_{.1}$	Nombre des 1 suivis par une autre valeur
$P(t)$	Matrice de probabilité de transition de la chaîne de Markov
$P_{ij}(t)$	Probabilités conditionnelles de transition entre l'état i à l'instant $t-1$ et l'état j à l'instant t
$r_t(i)$	Quantité de précipitations à générer, au site i , au jour t
SDI	Indicateur de dépendance spatiale
$u_t(i)$	Nombre aléatoire uniforme $[0, 1]$
$u(n,1)$	Vecteur de n nombres aléatoires indépendants et uniformément $[0, 1]$ distribués

$u_{T\max}(n,1)$	Vecteur de n nombres aléatoires indépendants et normalement distribués utilisés pour les températures maximales
$u_{T\min}(n,1)$	Vecteur de n nombres aléatoires indépendants et normalement distribués utilisés pour les températures minimales
$u_{Sr}(n,1)$	Vecteur de n nombres aléatoires indépendants et normalement distribués utilisés pour les radiations solaires
$v_i(i)$	Nombre aléatoire, uniforme $[0, 1]$
$V(n,1)$	Vecteur de n nombres aléatoires spatialement autocorrélés
$V_{T\max}(n,1)$	Vecteur de n nombres aléatoires spatialement autocorrélés utilisés pour les températures maximales
$V_{T\min}(n,1)$	Vecteur de n nombres aléatoires spatialement autocorrélés utilisés pour les températures minimales
$V_{Sr}(n,1)$	Vecteur de n nombres aléatoires spatialement autocorrélés utilisés pour les radiations solaires
$W(n,n)$	Matrice de poids
w_{ij}	Poids spatial entre les sites i et j
w_{\max}	Valeur propre positive maximale de la matrice $W(n,n)$
w_{\min}	Valeur propre négative maximale, en valeur absolue, de la matrice $W(n,n)$
$x_{i,j}^s$	Successeur de $D_{j(i)}$
$\chi_{p,k}(j)$	Matrice (3×1) comportant les résidus des températures maximales ($j = 1$), des températures minimales ($j = 2$) et des radiations solaires ($j = 3$) pour le jour k de l'année p
$X_{p,k}(j)$	Valeurs, au jour k de l'année p , des températures maximales ($j = 1$) et minimales ($j = 2$) et des radiations solaires ($j = 3$) pour le jour k de l'année p
$\bar{X}_k(j)$	Moyennes des variables $X_{p,k}(j)$ correspondant à l'état sec ou pluvieux du jour k
X	Variable à utiliser dans l'équation de l'autocorrélation spatiale
x_i	Valeur de la variable X à un site i
x_j	Valeur de la variable X à un site j
\bar{x}	Moyenne de la variable X sur les n sites
π_1	Probabilité stationnaire de précipitation
$\lambda_t(i)$	Paramètre de la fonction exponentielle, au site i , au jour t
$\varepsilon_{p,k}(j)$	Matrice (3×1) contenant des nombres aléatoires standards gaussiens indépendants $N[0,1]$ pour le jour k de l'année p

$\sigma_k(j)$	Écarts types des variables $X_{p,k}(j)$ correspondant à l'état sec ou pluvieux du jour k
γ	Coefficient du processus de moyenne mobile spatialisé
$\gamma_{T\max}$	Coefficient du processus de moyenne mobile spatialisé utilisé pour les températures maximales
$\gamma_{T\min}$	Coefficient du processus de moyenne mobile spatialisé utilisé pour les températures minimales
γ_{Sr}	Coefficient du processus de moyenne mobile spatialisé utilisé pour les radiations solaires
$\rho_{\chi_{i,0} \chi_{j,0}}$	Coefficient de corrélation croisée de lag 0 entre les résidus des variables i et j
$\rho_{\chi_{i,0} \chi_{j,-1}}$	Coefficient de corrélation croisée de lag 1 entre les résidus actuels de la variable i et les résidus précédents de la variable j
$\rho_{\chi_{i,0} \chi_{i,-1}}$	Coefficient d'autocorrélation de lag 1 des résidus de la variable i

INTRODUCTION

Contexte

Les données climatiques sont exploitées par une multitude de domaines, notamment, l'hydrologie, l'agriculture et récemment pour l'évaluation des impacts des changements climatiques. Ce dernier sujet est venu se placer au premier rang des priorités de la communauté scientifique. Beaucoup de questions concernant la nature des changements climatiques, leur évolution dans le futur, leurs conséquences probables sur les plans environnementaux, sociaux, et économiques, sont à l'étude. Il s'agit de mettre en place des stratégies efficaces pour s'adapter aux nouvelles conditions environnementales qui, d'après le groupe d'experts intergouvernemental sur l'évolution du climat (GIEC), vont se poursuivre pendant des siècles, même si l'on arriverait à diminuer les émissions des gaz à effet de serre. Ces conditions environnementales pourraient devenir encore plus alarmantes si les émissions de ces gaz évoluent à une cadence supérieure à celle observée actuellement.

Problématique

La gestion des ressources hydriques, les aménagements, la conception, la planification et la gestion des systèmes hydriques sont des activités qui nécessitent des banques de données climatiques. Analyser l'hydrologie des bassins versants, notamment en ce qui a trait aux événements extrêmes, nécessite de longues séries temporelles de ces données. Des séries temporelles futures tenant compte des changements climatiques sont requises pour évaluer l'impact de ces changements sur l'hydrologie des bassins versants. Or, les données observées sont souvent limitées et ne reflètent pas l'état futur du climat. Des générateurs de climat sont ainsi développés pour simuler des séries temporelles de longueur théoriquement infinie avec des propriétés statistiques semblables à celles des observations avec lesquelles ils ont été calibrés. Pour simuler des séries de données futures, les paramètres des générateurs de climat peuvent être modifiés en utilisant des scénarios de changements climatiques issus des MCG. Ces derniers simulent aussi exactement que possible le système climatique mondial

(atmosphère, océan). Vu la résolution spatiale limitée des MCG, les générateurs de climat peuvent être également utilisés comme une technique statistique de mise à l'échelle nécessaire pour adapter les données, issues de ces modèles, à une échelle régionale et locale (Wilks, 1999a).

L'intérêt grandissant envers les générateurs de climat a donné naissance au cours des dernières décennies à différentes approches de génération de climat. Appréciés pour leur simplicité d'implantation et d'utilisation, les générateurs de climat uni-site, c'est-à-dire ne tenant pas compte des corrélations entre les stations météorologiques, ont été fréquemment utilisés sans toutefois se rendre compte de leurs effets sur les applications pour lesquelles ils ont été conçus. Il est notoire que négliger la cohérence spatiale entre les données météorologiques simulées, alors qu'une forte dépendance spatiale est exhibée réellement sur le bassin versant, peut conduire à des résultats irréalistes, notamment si les données simulées sont utilisées comme input dans des modèles hydrologiques ou agraires (Wilks et Wilby, 1999; Mehrotra *et al.*, 2005; Srikanthan et McMahon, 2001; Harmel *et al.*, 2002). Il s'agit donc d'une importante préoccupation qui a mobilisé un nombre de chercheurs à développer des approches multi-sites pour reproduire les corrélations observées entre les stations météorologiques. Cependant, ce gain d'attention attribué aux générateurs multi-sites s'est heurté avec la limitation, la complication et parfois l'infaisabilité des techniques de génération multi-sites, ce qui justifie le recours continu aux générateurs uni-site.

Objectif de la thèse

La présente thèse s'inscrit dans le cadre d'un projet sur le développement d'un générateur stochastique multi-site de climat pour évaluer les impacts des changements climatiques sur les crues extrêmes. Ce projet est mené conjointement par le laboratoire DRAME à l'École de technologie supérieure et le consortium Ouranos sur la climatologie régionale et l'adaptation aux changements climatiques.

Un générateur de climat basé sur le modèle WGEN de Richardson (1981) a été développé. Il permet de simuler les variables climatiques à des sites individuels. Cette thèse a donc pour objectif principal de développer une approche de génération multi-site des données climatiques pour une évaluation réaliste des impacts des changements climatiques sur l'hydrologie des bassins versants.

La réalisation de cet objectif principal dérive de la réalisation des objectifs secondaires suivants :

- Le développement d'une approche de génération multi-site pour les processus d'occurrences et de quantités de précipitations journalières.
- Le développement d'une approche de génération multi-site pour les autres variables climatiques comme les températures maximales, minimales et les radiations solaires.
- La réalisation d'une étude de modélisation hydrologique spatialisée pour démontrer la supériorité et la performance de l'approche de génération multi-site des données climatiques, notamment pour évaluer avec plus de précision et de fiabilité les fréquences et magnitudes des crues extrêmes.

Ordre de la thèse

Ce travail doctoral est présenté dans cette thèse par le biais de cinq chapitres :

Chapitre 1 : Une revue bibliographique des différents travaux réalisés sur les générateurs de climat paramétrique et non paramétriques, uni-site et multi-sites, est présentée.

Chapitre 2 : Présentation des différents processus de génération uni-site, adaptés à partir du générateur WGEN de Richardson (1981), et de l'approche de régionalisation proposée pour reproduire la dépendance spatiale entre les données climatiques générées.

Chapitre 3 : Présentation de la première originalité de ce travail doctoral. Il s'agit du développement d'une nouvelle approche de génération multi-site des données de précipitation journalières basée sur le concept de l'autocorrélation spatiale. Ce chapitre est sous forme d'un article intitulé « Stochastic Multi-Site Generation of Daily Precipitation Data using Spatial Autocorrelation », publié dans le Journal of Hydrometeorology (Khalili *et al.* 2007).

Chapitre 4 : Présentation du deuxième aspect original de cette thèse. C'est le développement d'une nouvelle approche de génération multi-site des données de températures et de radiations solaires basée également sur le concept de l'autocorrélation spatiale. Une nouvelle approche est aussi développée pour reproduire l'intermittence spatiale des quantités de précipitations. Ce chapitre est le deuxième article de cette thèse. Il est intitulé « Stochastic Multi-site Generation of Daily Weather Data » et est sous évaluation au Journal of Hydrometeorology (Khalili *et al.* 2008a).

Chapitre 5 : C'est le troisième article réalisé dans le cadre de cette thèse. Il montre la supériorité du générateur de climat multi-site (Khalili *et al.* 2007, 2008a) pour la modélisation hydrologique en comparaison avec le générateur uni-site (Richardson, 1981). Cet article (Khalili *et al.* 2008b) dont le titre est « Effectiveness of Multi-Site Weather Generator for Hydrological Modelling: Comparative Study with Uni-Site Weather Generator using Distributed Hydrological Model » est sous évaluation au Journal of Hydrology.

La conclusion de cette thèse est présentée à la fin ainsi que des recommandations pour les recherches futures.

CHAPITRE 1

REVUE BIBLIOGRAPHIQUE

1.1 Introduction

Les générateurs de climat sont des modèles statistiques qui permettent de générer des données climatiques statistiquement similaires à des données observées. La demande incessante des données climatiques pour des études hydrologiques, agraires ou environnementales est derrière leur origine et évolution. Outre la modélisation du climat, ces outils statistiques peuvent être utilisés pour la simulation des données climatiques manquantes comme les radiations solaires qui sont souvent fournies par un nombre limité de stations (Richardson et Wright, 1984; Hanson *et al.*, 1994). Les générateurs de climat permettent la simulation des données manquantes, à un site donné, en interpolant leurs paramètres entre deux sites où les données sont disponibles (Dennett *et al.* 1983). Ils permettent également la génération des données journalières manquantes en se basant sur les statistiques mensuelles ou saisonnières disponibles (Katz, 1996; Wilks, 1992). Les générateurs de climat peuvent aussi être utilisés comme une technique statistique de mise à l'échelle pour adapter les données issues des modèles de circulation générale, dont la résolution spatiale est limitée, à une échelle régionale et locale (Wilks, 1999a). Avec les changements climatiques, les paramètres des générateurs de climat peuvent être modifiés afin de simuler des séries temporelles de climat futur.

L'émergence des générateurs de climat, abstraction faite de toute dépendance spatiale entre les données climatiques, a débuté vers les années 50 avec le développement des premiers générateurs de précipitations (Williams, 1952; Longley, 1953). Depuis, des travaux se sont multipliés pour la génération de cette variable jugée aussi bien importante que difficile à modéliser étant donné son caractère variable et asymétrique (exemple : Gabriel et Neuman, 1962; Bailey, 1964; Todorovic et Woolhiser, 1975; Buishand, 1978; Woolhiser et Pegram, 1979; Richardson, 1981; Stern et Coe, 1984; et Wilks, 1989, 1992). Les autres variables

climatiques, initialement marginalisées, n'avaient pas suscité autant d'attention. La littérature dévoile quelques études particulièrement circonscrites autour des températures maximales, minimales et des radiations solaires (Joseph, 1973 ; Nicks, 1975 ; Gutry-Korycka *et al.* 1994 ; Richardson, 1981 ; Young, 1994).

L'utilisation des données climatiques générées pour des projets décisionnels requiert une représentation aussi fidèle que possible des propriétés statistiques observées réellement. La dépendance spatiale est l'une des caractéristiques statistiques endémiques aux données climatiques qu'il importe de conserver lors de la génération de ces données. L'effet crucial que la dépendance spatiale peut avoir sur les études requérant des données climatiques, comme la gestion des ressources hydriques et, plus récemment, l'évaluation des impacts des changements climatiques sur les régimes hydrologiques des bassins versants, a fait surgir l'idée de développer des générateurs de climat multi-sites. La régionalisation des générateurs de climat s'est ainsi pointée au cours des dernières années sans toutefois s'imposer sur les techniques uni-site continuellement utilisées.

Que ce soit pour les générateurs multi-sites ou uni-site, l'état des connaissances mis au point a montré deux voix distinctes pour l'élaboration des modèles de génération des données climatiques. Il s'agit de l'approche paramétrique pour laquelle la modélisation nécessite l'établissement et l'estimation de paramètres ainsi que l'approche non paramétrique, moins restrictive, pour laquelle aucune hypothèse ou paramètres ne sont requis.

1.2 Revue bibliographique sur les générateurs de climat

1.2.1 Générateurs de climat paramétriques uni-site

Les travaux réalisés depuis de nombreuses décades ont contribué dans une grande mesure à mettre en œuvre les générateurs de climat utilisés actuellement. Le développement de ces générateurs a été amorcé par la génération des occurrences de précipitations, représentés par

des états secs et pluvieux. Ces états ont souvent tendance à se regrouper dans un processus d'occurrence manifestant ainsi une persistance définie par la dépendance statistique, à travers le temps, entre les valeurs d'une série (Wilks, 1995). Alors que certains modèles simulent des périodes de jours successifs secs et pluvieux (Williams, 1952; Longley, 1953; Green, 1964; Buishand, 1977; Roldan et Woolhiser, 1982; Racsko *et al.* 1991; Semenov et Barrow, 1997; Semenov *et al.*, 1998), d'autres préconisent la simulation des occurrences, jour par jour, par le modèle de chaîne de Markov. En optant pour cette dernière approche, des études ont démontré la bonne performance du modèle de chaîne de Markov de premier ordre (Gabriel et Newmann, 1962; Bailey, 1964; Smith et Schreiber, 1973 ; Haan *et al.*, 1976 ; Selvalingam et Miura, 1978 ; Richardson, 1981 ; Stern et Coe, 1984 ; Harrison et Waylen, 2000). D'autres auteurs ont par ailleurs préconisé un ordre supérieur pour améliorer la reproduction des longueurs, parfois limitées, des périodes sèches ou humides (Chin, 1977 ; Gates et Tong, 1976 ; Singh *et al.*, 1981 ; Lana et Burgueno, 1998 ; Caron, 2006). Le critère AIC (Akaike Information Criterion) (Akaike, 1974) ou bien le critère BIC (Bayesian Information Criterion) (Schwarz, 1978) sont souvent utilisés pour déterminer l'ordre optimal de la chaîne de Markov à utiliser. Le choix de cet ordre sera, toutefois, relié à l'espace et au temps étudiés (Srikanthan et McMahon, 2001).

Des modèles ont été élaborés pour la génération des quantités de précipitations. Il faut citer le modèle stochastique précurseur de Todorovic et Woolhiser (1975) pour la détermination de la fonction de distribution de la quantité totale de précipitations et la précipitation journalière extrême dans une période de n jours. La quantité journalière de précipitations a été modélisée comme une variable aléatoire de distribution exponentielle. Katz (1977) a présenté, par la suite, un modèle pour les quantités journalières de précipitations employant la chaîne de Markov. La quantité de précipitations produite à un jour pluvieux est choisie parmi deux distributions conditionnelles selon le caractère sec ou pluvieux du jour précédent. Le modèle élaboré par Todorovic et Woolhiser (1975) est un cas particulier du modèle de Katz (1977), en prenant les deux distributions conditionnelles identiques et exponentielles. Avec un nombre plus élevé de paramètres, la fonction gamma (Buishand, 1978 ; Coe et Stern, 1982 ; Richardson, 1981) et la fonction exponentielle mixte (Woolhiser et Pegram,

1979 ; Woolhiser et Roldan, 1982, 1986) ont semblé, ensuite, mieux représenter la distribution des quantités de précipitations. Par ailleurs, Semenov et Barrow (1997) et Semenov et al. (1998) ont utilisé des distributions semi-empiriques.

L'utilisation de la circulation atmosphérique dans des modèles climatiques a eu un effet positif dans de nombreuses études. Les générateurs de climat qui tiennent compte de ce type d'information à grande échelle peuvent servir comme technique de mise à l'échelle permettant d'obtenir des scénarios de climat pour évaluer les impacts des changements climatiques (Hay *et al.*, 1991; Wilks, 1992; Semenov et Barrow, 1997; Bardossy et Plate, 1991; Bardossy, 1991; Corte-Real *et al.*, 1998, 1999a, 1999b).

En raison de leur caractère peu variable, comparativement aux précipitations, les autres variables climatiques, notamment les températures maximales, minimales et les radiations solaires, imposent moins de contraintes quant à leur modélisation. Au début, ces variables climatiques ont été générées séparément par des distributions indépendantes, en se basant sur les processus d'occurrences des précipitations (Jones *et al.* 1972 ; Bruhn *et al.* 1980 ; Lane et Nearing, 1989). Par conséquent, les fortes autocorrélations et corrélations croisées, qui caractérisent ces variables, ont été ignorées. Nicks et Harp (1980) ont utilisé un modèle autorégressif d'ordre 1, indépendamment pour chaque variable, pour décrire le caractère autocorrélé de celle-ci. Les corrélations croisées entre les différentes variables sont restées, toutefois, négligées. Le modèle de Richardson (1981) (voir section 2.1.3 du chapitre 2 et section 4.3 du chapitre 4) est venu préserver ces caractéristiques statistiques en utilisant le modèle « weakly stationary generating process » de Matalas (1967). Considérés normalement distribués, les résidus des variables, obtenus en soustrayant les moyennes et en divisant par les écarts types, ont été simulés par un modèle autorégressif multivarié d'ordre 1. Les paramètres de ce modèle sont des matrices, fonctions des matrices de corrélations de lag 0 et 1 faisant intervenir les trois variables. Les paramètres de ce modèle sont supposés être fixes pour toutes les stations et les saisons (Richardson et Wright, 1984). Cette supposition a été, cependant, invraisemblable pour certains auteurs qui ont opté pour le calcul des matrices

pour chaque site et pour chaque saison (Hayhoe, 1998). Par ailleurs, les données de vent et d'humidité, requis particulièrement par les modèles agraires, ont été prises en compte (Wallis et Griffiths, 1997 ; Parlange et Katz, 2000; Bruhn *et al.*, 1980; Richardson et Nicks, 1990).

Le premier générateur de climat uni-site, mis au point, qui tient compte des données de précipitations, de températures maximales, minimales et de radiations solaires, est le WGEN développé aux USA par Richardson (1981). Ce générateur utilise le modèle de chaîne de Markov et la fonction gamma pour la génération des précipitations ainsi que le modèle «weakly stationary generating process» (Matalas, 1967) pour la génération des températures maximales, minimales et des radiations solaires. Un autre générateur de climat appelé LARS-WG (Racsko *et al.*, 1991 ; Semenov et Barrow, 1997) a été, par la suite, développé par le «Long Ashton Research Station» en Grande Bretagne. Ce générateur utilise des distributions semi-empiriques pour la génération des données de précipitations et de radiations solaires. Les températures maximales et minimales sont générées par une fonction de distribution normale. La majorité des générateurs de climat subséquents sont fondés sur ces deux types de générateurs, notamment le WGEN auquel les chercheurs se réfèrent le plus souvent. Des approches différentes peuvent être adoptées pour la simulation des paramètres.

1.2.2 Générateurs de climat paramétriques multi-sites

Le passage aux modèles de génération multi-sites a été initialement entrepris par des modèles espace-temps (Bras et Rodriguez-Iturbe, 1976; Waymire *et al.*, 1984; Smith et Karr, 1985). Relier la circulation atmosphérique aux variables climatiques est une théorie qui était confirmée par Burger (1958) lors de son étude de la relation entre les patrons de circulation atmosphérique et les températures journalières moyennes, maximales et minimales, les quantités de précipitations et la nébulosité. Bardossy et Plate (1992) ont repris ce principe pour développer un modèle espace-temps, traitant conjointement les occurrences et les quantités de précipitations, en considérant les patrons de circulation atmosphérique comme facteur principal de conditionnement. Dans le cas où le patron de circulation atmosphérique s'est avéré persistant, le processus aléatoire qui décrit les fonctions aléatoires distribuées

normalement a été considéré comme un processus autorégressif multivarié d'ordre 1 (AR1). Si le patron de circulation atmosphérique changeait à un temps t , un autre modèle devrait être considéré. Bardossy et Plate (1992) ont noté que la recherche du lien entre les patrons de circulation et les précipitations doit être effectuée sur une base historique. L'utilisation de la circulation atmosphérique a été reprise par Bardossy (1995a) pour la simulation des précipitations journalières et des températures moyennes journalières. Bardossy (1995b) a ensuite utilisé les patrons de circulation atmosphérique, définis à partir des distributions observées de pressions d'air journalières, pour simuler les données de précipitations.

Par ailleurs, Hughes et Guttorp (1994a) ont décrit une large classe de modèles, appelés Nonhomogeneous Hidden Markov Models (NHMM), pour la génération des occurrences des précipitations. Ces modèles supposent un état de climat non défini explicitement, considéré non observé et décrit par une chaîne de Markov de premier ordre. Les états de climat représentent les patrons atmosphériques prédictifs du processus hydrologique et sont détectés par le NHMM qui fonctionne comme un classificateur automatique. D'après Hughes et Guttorp (1994a), le NHMM englobe le modèle de chaîne de Markov simple (Gabriel et Neuman, 1962), le modèle de Markov caché (Zucchini et Guttorp, 1991) et les modèles espace-temps (Bardossy et Plate, 1992; Hay *et al.*, 1991; Hughes *et al.*, 1993). Hughes et Guttorp (1994a) ont signalé que si les stations météorologiques sont très séparées les unes des autres, leur interdépendance peut être expliquée par l'état commun du climat. Cependant, si les stations sont proches les unes des autres, les caractéristiques géographiques et météorologiques locales induisent des corrélations, entre les stations, non explicables par le climat à grande échelle. Hughes et Guttorp (1994b) ont, de ce fait, proposé un modèle de données multivariées d'occurrences des précipitations qui tient compte de la dépendance spatiale entre les stations. C'est une généralisation du modèle de Cressie (1991) pour des données binaires multivariées. Ce modèle a été incorporé dans le NHMM pour relier les mesures atmosphériques synoptiques aux données de précipitations locales. L'extension du NHMM pour la simulation des quantités de précipitations a été réalisée par Charles *et al.* (1999).

La mise en application des modèles espace-temps et NHMM demeure compliquée sans, toutefois, conduire à une reproduction adéquate de la cohérence spatiale exhibée par les données observées (Wilks, 1999b; Qian *et al.*, 2002). La complexité de ces modèles et le nombre élevé de paramètres mis en jeu laissent les utilisateurs réticents vis-à-vis les générateurs de climat multi-sites. Il faut mentionner que l'élaboration de ce type de générateur nécessite, avant tout, un questionnement au niveau de sa performance et de sa simplicité d'implantation sans oublier son côté parcimonieux.

Wilks (1998) a proposé un modèle stochastique multi-site basé sur le générateur de Richardson (1981). Il s'agit d'une méthode de régionalisation faisant appel à des nombres aléatoires standards gaussiens, caractérisés par une indépendance temporelle et une corrélation spatiale, permettant de reproduire les corrélations observées entre les séries synthétiques de précipitations. Ce modèle multi-site a donné des résultats adéquats par rapport à la reproduction des corrélations mensuelles observées. Ce qu'on lui reproche, toutefois, c'est sa structure encombrante faisant intervenir un nombre élevé de paramètres reliés aux corrélations entre chaque couple de stations utilisées. Un nombre élevé de relations doit être établi entre les corrélations des nombres aléatoires et celles de chaque processus de précipitations (occurrences et quantités), pour chaque couple de stations et pour chaque mois, afin d'obtenir les corrélations adéquates des nombres aléatoires qui vont reproduire les corrélations observées entre les processus de précipitations. Ce nombre de relations est égal à $12 \times k(k-1)$ pour un réseau de k stations. L'obtention des nombres aléatoires standards gaussiens, spatialement corrélés, est réalisée à partir d'une distribution normale multivariée de moyenne zéro et de variance-covariance représentée par la matrice qui comportent les corrélations mensuelles entre ces nombres aléatoires. Cette matrice n'est, cependant, pas toujours définie positive (i.e. une ou plusieurs valeurs propres sont négatives), ce qui empêche l'obtention des nombres aléatoires avec la décomposition de Cholesky. Pour pallier ce problème, des fonctions de distances entre les stations ont été suggérées pour obtenir les corrélations mensuelles entre ces nombres aléatoires. La fonction exponentielle mixte, utilisée pour réaliser la propriété d'intermittence spatiale observée entre les quantités de précipitations, n'a pas donné des résultats satisfaisants.

L'approche de Wilks (1998) a été appliquée aux données de précipitations du bassin versant de la rivière Châteauguay (Khalili *et al.* 2004a) ainsi qu'aux données de précipitations du bassin versant de la rivière Péribonca (Khalili *et al.* 2004b). Une comparaison entre cette approche et l'approche développée dans le cadre de cette thèse est effectuée. Les résultats de cette comparaison sont présentés dans le premier article, objet du chapitre III. Brissette *et al.* (2007) ont développé un algorithme de calcul pour simplifier l'application de cette méthode.

Par ailleurs, Wilks (1999b) a proposé un modèle multi-site pour la simulation des données journalières de températures maximales, minimales et de radiations solaires en se basant sur le modèle autorégressif d'ordre 1, proposé par Richardson (1981). En effet, une extension du modèle autorégressif a été effectuée en intégrant toutes les stations utilisées. La dimension du modèle autorégressif est ainsi passée de 3 à $3k$ pour un réseau de k stations. Dans ce cas, les paramètres matriciels, fonctions des matrices de corrélations de lag 0 et 1 faisant intervenir les trois variables, sont de dimensions $3k \times 3k$. Ces dimensions augmentent davantage au fur et à mesure que le réseau de stations s'élargit ou bien qu'une autre variable climatique est prise en compte, ce qui complique la solution du modèle. Des ajustements pour les matrices de corrélations doivent être effectués pour la faisabilité des calculs.

Qian et al. (2002) ont noté que les séries de précipitations journalières, simulées séparément à différents sites, peuvent également montrer une corrélation si les patrons de circulation journalière ont été considérés. Incertains de la consistance de ces corrélations, notamment pour les quantités de précipitations, ils ont souligné la nécessité d'une régionalisation des générateurs de climat, conditionnés ou non par les patrons de circulation, pour reproduire adéquatement la corrélation spatiale des stations. Qian et al. (2002) ont repris le modèle de Wilks (1998, 1999b) en intégrant les patrons de circulation journalière.

1.2.3 Générateurs de climat non paramétriques uni-site

Les générateurs de climat non paramétriques sont basés, principalement, sur la méthode du rééchantillonnage ou le « bootstrap ». Cette technique est reconnue comme non paramétrique

vu qu'aucune hypothèse n'est préalablement imposée, comme la forme de la distribution des données. Contrairement aux méthodes paramétriques, la simulation des précipitations et des autres variables climatiques se fait simultanément.

D'après Lall et Sharma (1996), le bootstrap (Efron, 1979; Efron et Tibishirani, 1993) est une technique qui prescrit une stratégie de rééchantillonnage des données en utilisant le mécanisme aléatoire qui a généré ces données. À l'origine, le bootstrap supposait que les données soient indépendantes. Cette hypothèse a été rectifiée, par la suite, vu le caractère autocorrélé des séries météorologiques ou hydrologiques. De ce fait, trois approches ont été développées pour introduire cette autocorrélation. Ces approches sont présentées par Brandsma et Buishand (1997). La première est le modèle de Markov utilisant le rééchantillonnage à partir des jours représentant le même état que le jour d'intérêt. Cette méthode a été utilisée par Yakowitz (1979) pour la simulation des débits journaliers. Hughes *et al.* (1993) et Conway *et al.* (1996) l'ont reprise pour la génération des précipitations journalières, conditionnée par la circulation atmosphérique. La deuxième méthode est le rééchantillonnage à partir des analogues ou proches voisins. Utilisée originellement pour les prévisions climatiques, la méthode des analogues a été appliquée ensuite par Zorita *et al.* (1995) pour une génération multi-site des précipitations journalières. Quant à l'approche des proches voisins, elle a été employée par Lall et Sharma (1996) pour la simulation des séries hydrologiques et par Rajagopalan et Lall (1995) pour la génération de données climatiques journalières multivariées. La troisième méthode de bootstrap est celle par blocs mobiles (the moving blocks bootstrap) utilisée par Vogel et Shallcross (1996) pour la génération des débits annuels. Une forme relativement nouvelle de bootstrap par blocs mobiles a été développée par Wilks (1997) pour la simulation de séries climatiques univariées.

La méthode des proches voisins est la plus répandue pour la génération non paramétrique. Young (1994) a utilisé cette méthode pour la simulation des données journalières des précipitations et des températures maximales et minimales. Une approche, théoriquement semblable à celle de Young (1994) mais différente en pratique, a été présentée par Rajagopalan et Lall (1999) pour la simulation des précipitations, radiations solaires,

températures maximales et minimales, de la température moyenne du point de rosée et de la vitesse moyenne du vent. Cette méthode est basée sur le rééchantillonnage à partir des successeurs des k proches voisins historiques d'un vecteur d'état actuel défini à priori. Au lieu de prélever de façon uniforme à partir des successeurs des k proches voisins, les auteurs ont utilisé le prélèvement discret de Kernel qui est monotone décroissant. Les variables considérées ont été standardisées en soustrayant la moyenne et en divisant par l'écart-type. Pour comprendre davantage cette approche de simulation, Rajagopalan et Lall (1999) décrivent ses étapes comme suit :

- Identifier la composition du vecteur d'état D_i de dimension d : Rajagopalan et Lall (1999) ont choisi de prendre, comme vecteur d'état, le vecteur composé des six variables du jour précédent, mentionnées ci-dessus,.
- Déterminer les k proches voisins du vecteur d'état actuel, noté D_i , parmi les vecteurs d'état historiques, par la distance euclidienne pondérée.
- Soit $J_{i,k}$ l'ensemble ordonné des k proches voisins et $D_{j(i)}$ le $j^{\text{ème}}$ proche voisin de D_i . $x_{i,j}^s$ est le successeur de $D_{j(i)}$.
- Définir le Kernel discret $k[j(i)]$ pour prélever un des $x_{i,j}^s$

$$k[j(i)] = \frac{1/j}{\sum_{j=1}^k 1/j} \quad (1.1)$$

avec $k[j(i)]$ est la probabilité de prélever $x_{i,j}^s$.

- Utiliser la fonction de densité de probabilité discrète $k[j(i)]$, prélever un $x_{i,j}^s$, mettre à jour le vecteur d'état et procéder à l'étape 2 pour d'autres simulations.

Brandsma et Buishand (1997) avaient adopté la même approche de Rajagopalan et Lall (1999) mais en intégrant des indices de circulation atmosphérique. En outre, deux types de prélèvement ont été considérés. Le premier est incondtionnel pour lequel le vecteur d'état

D_t contient les variables qui caractérisent l'état du climat au jour $t-1$ comme la précipitation journalière, la température et les indices de circulation atmosphérique. Le deuxième est conditionnel intégrant, dans le vecteur d'état D_t , les indices de circulation du jour t . Par conséquent, dans le cas de la simulation inconditionnelle, le successeur du jour prélevé, à partir des k proches voisins historiques, devient le nouveau jour généré, alors que dans le cas de la simulation conditionnelle, la précipitation et la température du jour à générer sont égales aux valeurs du jour prélevé lui même.

1.2.4 Générateurs de climat non paramétriques multi-sites

La régionalisation des modèles de génération non paramétriques a été réalisée par Buishand et Brandsma (2001). Ils ont développé une méthode pour la simulation multi-site des précipitations et des températures, pour 25 stations situées sur la partie allemande du bassin du Rhin, par prélèvement à partir des k proches voisins. Cette étude est une extension des méthodes utilisées par Lall et Sharma (1996) et Rajagopalan et Lall (1999). Pour reproduire la dépendance spatiale, les 25 observations de précipitations et les 25 observations de températures doivent toutes être intégrées dans le vecteur d'état D_t et ceci pour chaque jour. Devant ce calcul ardu qui peut résulter en une différence importante au niveau des précipitations locales, ces observations ont été représentées par un nombre faible de statistiques, dites sommaires, comme les moyennes arithmétiques, sur les 25 stations, des précipitations et des températures standardisées. Cependant, à cause de la variation spatiale importante des précipitations, la moyenne arithmétique des précipitations n'a expliqué que 53% de la variance journalière de cette variable. D'autres variables ont alors été incorporées comme les moyennes journalières des valeurs de précipitations standardisées, calculées sur 5 régions spécifiées sur le bassin, ou bien la fraction des stations avec des précipitations supérieures à un certain seuil.

Ainsi, le vecteur d'état D_t , qui comporte la moyenne des températures sur les 25 stations, doit intégrer les différentes statistiques concernant les précipitations. Dans le cas où les

indices de circulation ont été considérés, ceux du jour $t - 1$ ont été inclus dans le vecteur D_t afin de les simuler simultanément avec les températures et les précipitations. Pour la simulation conditionnelle, ce sont les indices du jour t qui ont été intégrés dans le vecteur d'état. Dans ce cas, le prélèvement a été limité aux jours avec une circulation similaire à celle du jour t .

Buishand et Brandsma (2001) ont conclu que puisque la méthode de simulation est basée sur le prélèvement à partir des données historiques, la dépendance entre les variables, à des sites différents, est automatiquement préservée. Cependant, des problèmes liés au principe même de l'approche ont été révélés :

« The method does not produce new daily precipitation and temperature fields but merely reshuffles historical days to form realistic new sequences of those fields. In the case of missing values, resampling of multivariate data may become quite problematic [...] Potential users of nearest-neighbor resampling should be aware of the consequences of repetitions of segments of the historical record in the simulated sequences and that the simulated daily rainfall amounts cannot be larger than those observed. The former can be mitigated at the cost of a poorer reproduction of temporal dependence, in particular that of the daily temperatures. The extension of the method to simulate sequences of subdaily values is not straightforward because of the stronger temporal dependence. ». Buishand et Brandsma (2001).

1.3 Conclusion

La revue bibliographique précédente a permis de souligner les principales caractéristiques de chaque gamme de modèles. Des lacunes ont été révélées notamment pour les générateurs multi-sites pour mieux appréhender les objectifs du nouveau générateur multi-site développé dans cette thèse. Bien qu'il s'agisse d'une approche relativement simple, sans aucune hypothèse a priori et aucune estimation des paramètres, l'approche non paramétrique semble inappropriée pour étudier les changements climatiques à cause de la répétition de segments de données historiques dans les séquences simulées. La sous-estimation des autocorrélations et des variances des valeurs journalières est une conséquence évidente de cette approche. En revanche, les modèles paramétriques sont aptes pour reproduire de nouvelles séries de

données et peuvent, convenablement, être modifiés pour tenir compte des changements climatiques. Le modèle multi-site de Wilks (1998), fondé sur la théorie de Richardson (1981), paraît le mieux coté mais demeure, toutefois, loin de la performance attendue. Tel que mentionné ci-dessus, des lacunes ont été décelées pour ce modèle. Les plus critiques sont celles concernant le nombre élevé de paramètres à prendre en compte, qui augmente au fur et à mesure que le réseau de stations augmente et les matrices de corrélations qui doivent être définies positives pour la faisabilité des calculs. En outre, pour les autres variables climatiques, l'extension du modèle autorégressif, pour intégrer toutes les stations utilisées, amplifie les matrices de corrélations ce qui empêche, ou du moins complique, la solution du modèle.

Le défi à relever par ce travail doctoral est donc de développer une nouvelle approche de régionalisation, efficace et pratique, qui permet de pallier les insuffisances rencontrées avec les autres générateurs multi-sites. Rappelons que cette thèse se situe dans le cadre d'un projet sur le développement et la régionalisation d'un générateur de climat basé sur l'approche de Richardson (1981).

Une description détaillée des processus de génération uni-site et de l'approche de régionalisation est fournie au chapitre suivant. L'application de l'approche de génération multi-site aux données de précipitations est présentée au chapitre 3. Le chapitre 4 présente l'approche de génération multi-site pour les autres variables climatiques. Le chapitre 5 fait état d'une étude de modélisation hydrologique réalisée pour tester la performance du générateur multi-site développé.

CHAPITRE 2

DESCRIPTION GÉNÉRALE DES PROCESSUS DE GÉNÉRATION UNI-SITE ET DE L'APPROCHE DE RÉGIONALISATION

2.1 Description des processus de génération uni-site

2.1.1 Génération des occurrences de précipitations

Le générateur multi-site développé dans la présente thèse se base sur le modèle uni-site WGEN de Richardson (1981). Comme la plupart des générateurs de climat paramétriques, il procède par la simulation des occurrences de précipitations, puis la simulation des quantités de précipitations et ensuite la simulation des autres variables météorologiques comme les températures maximales, minimales et les radiations solaires.

Les états d'occurrence des précipitations peuvent être représentés par deux valeurs. La valeur 0 pour l'état sec et la valeur 1 pour l'état pluvieux. Réputé supérieur pour la simulation des occurrences de précipitations, le modèle de chaîne de Markov a été considérablement utilisé par des générateurs de climat y compris WGEN. La chaîne de Markov est une théorie qui peut être appliquée pour modéliser les processus à caractère discret, comme les processus météorologiques ou hydrologiques. Elle peut être définie comme suit (Maidment, 1993):

Soit $X(t)$ un processus discret qui peut prendre les valeurs x_t avec $t = 0, 1, \dots$

La probabilité de prendre la valeur x_t à l'instant t en tenant compte de toutes les valeurs précédentes ou bien de l'historique du processus est :

$$P[X(t) = x_t | X(0) = x_0, X(1) = x_1, \dots, X(t-1) = x_{t-1}] \quad (2.1)$$

La chaîne de Markov simple ou chaîne de Markov de premier ordre est un processus dont la probabilité ci-dessus tient compte uniquement du résultat à l'instant précédent.

$$P[X(t) = x_t | X(t-1) = x_{t-1}] \quad (2.2)$$

Si l'on désire modéliser le processus régissant les précipitations journalières, deux états peuvent être considérés; $j = 0$ pour le jour sec sans précipitation et $j = 1$ pour le jour pluvieux. Ces états sont les valeurs x_t qui caractérisent le processus.

La chaîne de Markov de premier ordre est définie par une matrice appelée matrice de probabilité de transition $P(t)$. Cette dernière est une matrice carrée dont les éléments sont les probabilités conditionnelles de transition $P_{ij}(t)$ entre l'état i à l'instant $t-1$ et l'état j à l'instant t données par :

$$P_{ij}(t) = P[X(t) = j | X(t-1) = i] \quad (2.3)$$

Pour une chaîne de Markov de premier ordre et deux états, il existe 4 probabilités conditionnelles de transition, mais il suffit d'en connaître deux pour en déduire les deux autres (Wilks, 1995) :

$$\begin{aligned} p_{00} &= \Pr\{x_{t+1} = 0 | x_t = 0\} \\ p_{01} &= \Pr\{x_{t+1} = 1 | x_t = 0\} \\ p_{10} &= \Pr\{x_{t+1} = 0 | x_t = 1\} \\ p_{11} &= \Pr\{x_{t+1} = 1 | x_t = 1\} \end{aligned} \quad (2.4)$$

avec $p_{00} + p_{01} = 1$ et $p_{10} + p_{11} = 1$.

Wilks (1995) note que pour les quantités qui présentent une autocorrélation: $p_{10} < p_{00}$ et $p_{01} < p_{11}$. La chaîne de Markov est considérée stationnaire ou homogène si les probabilités de transition sont indépendantes du temps.

L'estimation des probabilités de transition revient au calcul des fréquences relatives conditionnelles (Wilks, 1995).

$$\begin{aligned}
p_{01} &= \frac{\text{nombre des 1 suivant les 0}}{\text{total des 0}} = \frac{n_{01}}{n_{0.}} \\
\text{et} \\
p_{11} &= \frac{\text{nombre des 1 suivant les 1}}{\text{total des 1}} = \frac{n_{11}}{n_{1.}}
\end{aligned} \tag{2.5}$$

avec

n_{01} est le nombre des 0 suivis par 1 ;

n_{11} est le nombre des 1 suivis par 1 ;

$n_{0.}$ est le nombre des 0 suivis par une autre valeur, $n_{0.} = n_{00} + n_{01}$;

$n_{1.}$ est le nombre des 1 suivis par une autre valeur, $n_{1.} = n_{10} + n_{11}$.

Wilks (1995) donne une description détaillée de la procédure de simulation des processus d'occurrences des précipitations. En effet, un générateur de nombre aléatoire uniforme $[0, 1]$ permet de générer des nombres aléatoires avec une probabilité égale. Un premier nombre généré sera comparé avec la probabilité stationnaire de précipitation π_1 définie par :

$$\pi_1 = \frac{p_{01}}{1 + p_{01} - p_{11}} \tag{2.6}$$

Si ce nombre est inférieur à π_1 , $x_1 = 1$ et le premier jour simulé sera humide. Sinon, $x_1 = 0$ et le premier jour simulé sera sec. D'autres nombres aléatoires seront ensuite générés pour simuler les jours suivants. Dans ce cas, chaque nouveau nombre aléatoire est comparé à la probabilité de transition appropriée selon l'état pluvieux ou sec du jour précédent. En effet, la probabilité de transition p_{01} sera utilisée pour générer x_{t+1} si $x_t = 0$ et p_{11} sera utilisée si $x_t = 1$. Si le nombre aléatoire généré est inférieur à la probabilité de transition, le jour simulé sera pluvieux $x_{t+1} = 1$; sinon, le jour simulé sera sec $x_{t+1} = 0$.

En partant du principe de Richardson (1981) pour la détermination des probabilités de transition, le générateur de climat uni-site, développé au laboratoire DRAME, calcule les probabilités de transition à partir des données observées sur 26 périodes de 14 jours. Un lissage par séries de Fourier est ensuite appliqué pour décrire les valeurs journalières de ces probabilités sur toute l'année. Il faut noter, que certaines modifications ont été apportées à ce générateur (Caron, 2006), ce qui a donné naissance à une nouvelle version appelée WeaGETS. Pour cette thèse, la première version utilisant le modèle de chaîne de Markov de premier ordre est utilisé mais la méthode de régionalisation développée peut être, convenablement, adaptée à la nouvelle version utilisant un ordre plus grand.

2.1.2 Génération des quantités de précipitations

Les distributions des quantités de précipitations présentent souvent une forte asymétrie à droite à cause des fortes fréquences des faibles quantités, ce qui les rend difficile à modéliser, notamment les précipitations extrêmes. Tel que mentionné au chapitre 1, la distribution exponentielle (Todorovic et Woolhiser, 1975) a été originellement utilisée pour modéliser les quantités de précipitations. Elle a été reprise dans d'autres études comme Wilby (1994). Buishand (1977, 1978), Katz (1977), Stern et Coe (1984), Thom (1958) et Wilks (1989, 1992) ont, par contre, opté pour la distribution gamma à deux paramètres qui fournit un meilleur ajustement aux quantités de précipitations. Développée par Woolhiser et Pegram (1979), la distribution exponentielle mixte à trois paramètres s'est avérée la plus performante (Woolhiser et Roldan, 1982; Foufoula-Georgiou et Lettenmaier, 1987; Wilks, 1998). Le lecteur est invité à se référer à Wilks et Wilby (1999) pour plus d'information sur ces types de distributions.

Le but de cette thèse n'est pas de chercher la distribution qui offre le meilleur ajustement aux quantités de précipitations, mais plutôt de chercher à reproduire la dépendance et l'intermittence spatiales qui caractérisent cette variable. En se basant sur le modèle de Richardson (1981), les fonctions exponentielle et gamma peuvent être aussi bien utilisées par l'approche multi-site élaborée dans le cadre de cette thèse. Une nouvelle approche de

modélisation des quantités de précipitations a, toutefois, été développée pour reproduire l'intermittence spatiale exhibée par les précipitations observées (voir section 4.2.2 du chapitre 4).

Le principe de génération uni-site des quantités de précipitations est basé sur l'inversion de la fonction de densité cumulative de la fonction de distribution choisie. Dans le cas d'une fonction exponentielle simple, cette fonction de densité cumulative est sous forme :

$$F[r_t(i)] = 1 - \exp(-\lambda_t(i) r_t(i)) \quad (2.7)$$

avec

$r_t(i)$ est la quantité de précipitations à générer, au site i , au jour t ;

$\lambda_t(i)$ est le paramètre de la fonction exponentielle, au site i , au jour t .

La quantité de précipitations $r_t(i)$ est déduite en utilisant un nombre aléatoire $v_t(i)$, uniforme $[0, 1]$, et l'inverse de la fonction F ci-dessus, tel que :

$$r_t(i) = -\ln(1 - v_t(i)) / \lambda_t(i) \quad (2.8)$$

Comme pour le cas des probabilités de transition, le générateur de climat uni-site calcule le paramètre $\lambda_t(i)$, considéré égal à l'inverse de la moyenne de la distribution exponentielle, sur 26 périodes de 14 jours. Le lissage par séries de Fourier est utilisé pour déterminer les valeurs journalières de ce paramètre.

2.1.3 Génération des autres variables climatiques

Le modèle de Richardson (1981) était le premier à avoir élargi la fonction des générateurs de climat pour simuler, convenablement, les autres variables météorologiques telles que les températures maximales, les températures minimales et les radiations solaires à un site individuel. Ces variables ont été considérées comme un processus stochastique multivarié dont les moyennes journalières et les écarts-types dépendent des états des jours, secs ou pluvieux, déterminés préalablement par les processus d'occurrences. Les séries chronologiques de chaque variable ont été transformées en séries de résidus en éliminant les moyennes et en divisant par les écarts-types. Ces statistiques ont été calculées séparément selon les états secs et pluvieux et ont été lissées sur toute l'année par des séries de Fourier.

Les séries des résidus caractérisées par des autocorrélations et des corrélations croisées ont été modélisées par un modèle autorégressif multivarié (weakly stationary generating process) suggéré par Matalas (1967).

$$\chi_{p,k}(j) = A \chi_{p,k-1}(j) + B \varepsilon_{p,k}(j) \quad (2.9)$$

$\chi_{p,k}(j)$ et $\chi_{p,k-1}(j)$ sont des matrices (3×1) contenant les résidus des températures maximales ($j = 1$), des températures minimales ($j = 2$) et des radiations solaires ($j = 3$) pour les jours k et $k - 1$ de l'année p ;

$\varepsilon_{p,k}(j)$ est une matrice (3×1) contenant des nombres aléatoires standards gaussiens indépendants $N[0,1]$ pour le jour k de l'année p ;

A et B sont des matrices (3×3) chacune, déduites de deux autres matrices M_0 et M_1 contenant les autocorrélations et les corrélations croisées de lag 0 et 1, respectivement, des résidus observés (voir section 4.3 du chapitre 4 pour plus de détails).

Les valeurs $X_{p,k}(j)$, au jour k de l'année p , des températures maximales et minimales et des radiations solaires sont, ensuite, obtenues en additionnant les moyennes aux résidus simulés et en multipliant par les écarts-types.

$$X_{p,k}(j) = (\chi_{p,k}(j) + \bar{X}_k(j)) \times \sigma_k(j) \quad (2.10)$$

avec $\bar{X}_k(j)$ et $\sigma_k(j)$ sont, respectivement, les moyennes et les écarts-types, des variables $X_{p,k}(j)$, correspondant à l'état sec ou pluvieux du jour k .

2.2 Approche de régionalisation : l'autocorrélation spatiale

Les données climatiques, dites aussi météorologiques, puisque seul l'attribut temporel fait la différence, sont des données descriptives (quantités des précipitations ou degrés de température) et géométriques (localisation des stations météorologiques). Ce double caractère fait de l'autocorrélation spatiale une propriété fondamentale des données climatiques et un outil privilégié de prise en compte de la dépendance spatiale dans le générateur de climat multi-site développé dans cette thèse.

2.2.1 Qu'est ce que l'autocorrélation spatiale ?

L'autocorrélation spatiale est l'analogue de l'autocorrélation temporelle, connue sous le vocable « autocorrélation » (serial autocorrelation), très utilisée en statistique pour modéliser le comportement autorégressif des séries chronologiques univariées. L'autocorrélation spatiale est donc la corrélation entre les valeurs, d'une même variable, spatialement distribuées. Autrement dit, c'est une statistique qui permet de mesurer la dépendance spatiale entre les valeurs, d'une même variable, en tenant compte de leur positionnement dans l'espace. Alors que l'autocorrélation temporelle peut provenir de la persistance temporelle d'un phénomène à un site particulier, l'autocorrélation spatiale, quant à elle, résulte de l'extension spatiale d'un phénomène sur plusieurs sites.

L'autocorrélation spatiale est souvent exprimée par le coefficient I de Moran (Moran, 1950; Cliff et Ord, 1981).

$$I = \frac{\sum_{i=1}^n (x_i - \bar{x}) \sum_{j=1}^n w_{ij} (x_j - \bar{x}) / \sum_{i=1}^n \sum_{j=1}^n w_{ij}}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2 / n} \sqrt{\sum_{i=1}^n (x_i - \bar{x})^2 / n}} \quad (2.11)$$

n est le nombre de sites;

x_i est la valeur de la variable X à étudier, à un site i ;

x_j est la valeur de la variable X à un site j ;

\bar{x} est la moyenne de la variable X sur les n sites;

w_{ij} est un poids spatial entre les sites i et j .

Le coefficient I de Moran varie, généralement, entre -1 et 1 avec une valeur moyenne de $-1/(n-1)$. Quand I est supérieur à cette valeur, l'autocorrélation spatiale est dite positive, ce qui signifie que les valeurs de la variable, géographiquement rapprochées, ont tendance à être similaires (Figure 2.1). En effet, les grandes valeurs ont tendance à être localisées près des grandes valeurs et les petites valeurs près des petites valeurs. Par contre, quand I est inférieur à la valeur moyenne, l'autocorrélation spatiale est dite négative, ce qui signifie que les valeurs de la variable, géographiquement rapprochées, ont tendance à être dissimilaires. Par conséquent, les grandes valeurs ont tendance à être localisées près des petites valeurs et vice versa. En outre, une valeur de I égale à la valeur moyenne signifie que les valeurs sont aléatoirement distribuées dans l'espace.

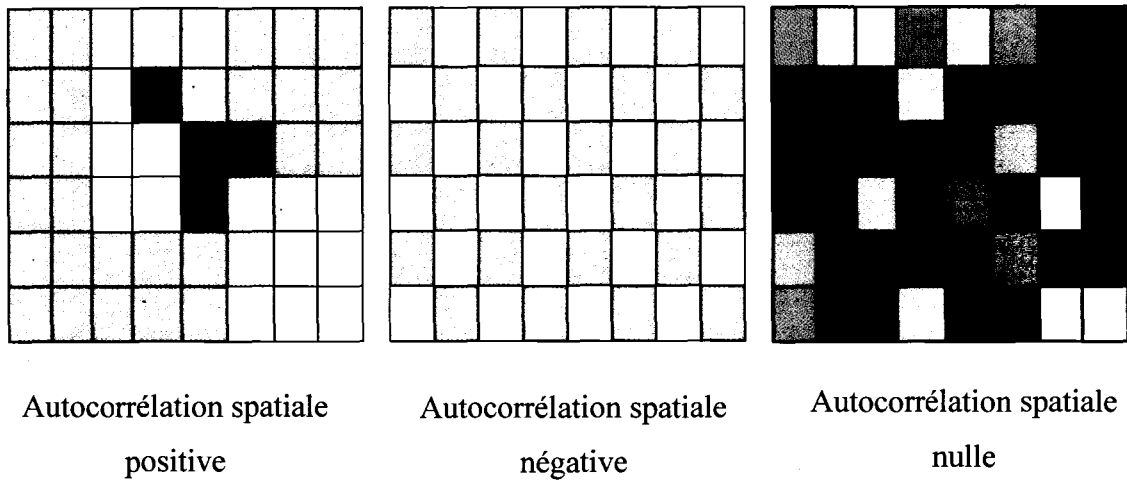


Figure 2.1 *Types d'autocorrélations spatiales*

L'autocorrélation spatiale peut être exprimée autrement par l'indice C de Geary (Odland, 1988; Griffith, 2003) :

$$C = \frac{n-1}{2 \sum_i \sum_j w_{ij}} \frac{\sum_i \sum_j w_{ij} (x_i - x_j)^2}{\sum_i (x_i - \bar{x})^2} \quad (2.12)$$

Cet indice est compris entre 0 et 2. Lorsqu'il est égal à 1, les valeurs sont considérées indépendantes. Lorsqu'il est inférieur à 1, l'autocorrélation spatiale est positive, tandis qu'elle est négative si cet indice est entre 1 et 2. D'après Cliff et Ord (1981), l'indice de Moran semble, statistiquement, plus avantageux par rapport à celui de Geary. Un avantage, qui paraît évident, est sa structure qui est presque similaire à celle du coefficient de corrélation standard. La différence réside dans l'utilisation d'une seule variable X , au lieu de deux variables X et Y , et des poids. Ses résultats sont aussi similaires aux résultats du coefficient de corrélation, ce qui facilite son interprétation. Pour cette thèse, le choix s'est porté sur l'indice de Moran pour les raisons préalablement citées.

Pour mesurer l'autocorrélation spatiale, il faut définir la matrice de poids, dite aussi matrice de voisinage. Il s'agit d'une matrice contenant les poids w_{ij} décrivant l'interaction spatiale entre chaque couple de localités. Différentes expressions de poids peuvent être trouvées dans la littérature (Odland 1988; Anselin et Bera, 1998; Anselin 1980; Murdoch *et al.* 1993). Par exemple, ces poids peuvent être exprimés en fonction des distances entre les sites, comme l'inverse de la distance au carrée ou bien en utilisant une distance d fixée à priori, tel que :

$$\begin{aligned} w_{ij} &= 1 & \text{si } d_{ij} < d \\ w_{ij} &= 0 & \text{sinon} \end{aligned} \quad (2.13)$$

Par ailleurs, les fonctions de pondération spatiale utilisées pour l'autocorrélation spatiale, ne représentent pas uniquement le poids par rapport à la distance. Odland (1988) a noté que ces fonctions peuvent, généralement, être considérées comme des moyens pour accommoder des hypothèses par rapport aux relations entre les sites. Ces hypothèses seront souvent de simples variantes de la première loi de géographie proposée par Tobler (1970) «everything is related to everything else, but near things are more related than distant things». Dans cette thèse, des interrelations plus réalistes se sont avérées nécessaires afin de conserver la dépendance réelle entre chaque couple de stations météorologiques (chapitres 3 et 4).

2.2.2 Pourquoi l'autocorrélation spatiale ?

L'autocorrélation spatiale intéressait, depuis longtemps, diverses disciplines, par exemple : l'écologie, l'archéologie, l'épidémiologie, la sociologie et la géologie (Odland, 1988). Récemment, l'autocorrélation spatiale était employée dans diverses études d'économétrie spatiale (Anselin et Bera, 1998; Le Gallo, 2000). Le facteur commun entre ces différents domaines est l'interaction spatiale pour laquelle des événements, à un endroit donné, sont reliés à ceux aux endroits voisins. Il s'agit, par exemple, du mouvement des individus, des capitaux, des espèces, des biens, des prix, des informations, des technologies, des accidents routiers, des crimes, etc.

La climatologie et la météorologie fondées sur le passage des fronts chauds et froids font aussi état de l'autocorrélation spatiale puisque le mouvement des masses d'air a tendance de se développer sur une région au complet au lieu d'une seule localité. Les sites sont ainsi affectés par le même (ou presque le même) phénomène météorologique, ce qui induit l'autocorrélation spatiale entre les données issues des sites d'observations.

Si l'autocorrélation spatiale est présente dans les données climatiques, de quelle façon doit être utilisée pour une génération multi-site? C'était la question que l'auteure s'est posée quand elle a été confrontée à cette notion statistique.

Il faut noter, d'abord, que l'autocorrélation spatiale a un effet attrayant pour ce projet vu qu'elle représente, par un seul nombre, l'interaction spatiale sur une région au complet. Contrairement aux corrélations standards (de Pearson) qui fournissent la dépendance entre deux variables X et Y , l'autocorrélation spatiale, quant à elle, mesure la dépendance entre les valeurs d'une même variable X , sous contrainte du positionnement de ces valeurs dans l'espace. Dans le cadre de cette thèse, cette variable X peut être soit des séries d'occurrences de précipitations, quantités de précipitations, températures maximales, températures minimales, radiations solaires ou, éventuellement, d'autres variables climatiques.

La réponse à la question ci-dessus était de procéder au niveau des approches de génération uni-site, présentées au début de ce chapitre, pour tenir compte de l'autocorrélation spatiale observée réellement entre les sites météorologiques. Si l'on doit spécifier un paramètre particulier pour faire intervenir cette autocorrélation spatiale, il va s'agir sans aucun doute des nombres aléatoires. Ces derniers interviennent dans la génération des occurrences de précipitations via la comparaison avec la probabilité de transition critique, dans la génération des quantités de précipitations à travers l'inverse de la fonction de densité cumulative de la fonction de distribution choisie et dans la génération des autres variables météorologiques par la composante aléatoire du modèle autorégressif. L'idée est de choisir des nombres aléatoires spatialement autocorrélés qui peuvent reproduire les autocorrélations spatiales observées sur le terrain d'étude.

Signalons qu'une intervention au niveau des nombres aléatoires du générateur WGEN a été effectuée par Wilks (1998) pour la génération multi-site des données de précipitations (voir section 1.2.2 de la revue bibliographique). Il a utilisé des nombres aléatoires spatialement corrélés pour reproduire les corrélations mensuelles observées entre chaque couple de stations météorologiques. Le lecteur aura compris que l'approche de Wilks (1998) considère les corrélations standards (corrélations de Pearson qui mesurent la relation entre deux variables X et Y) et traite chaque couple de stations séparément. Tel que mentionné avant, cette approche engendre un nombre élevé de paramètres qui croît au fur et à mesure que le réseau de stations augmente. L'autre lacune notée est la contrainte d'avoir des matrices de corrélations définies positives pour la faisabilité des calculs. Les autres variables météorologiques ont été modélisées, autrement, en augmentant la taille du modèle autorégressif pour intégrer tous les sites météorologiques existants. Cette méthode entrave la solution du problème surtout avec un nombre élevé de stations ou de variables climatiques.

Quoique conçue pour une description globale de la dépendance spatiale, l'autocorrélation spatiale a été ici utilisée avec succès de façon à reproduire les corrélations (de Pearson) entre les couples de stations météorologiques, via la matrice de poids. Il faut noter que le fait de reproduire l'autocorrélation spatiale observée, avec une description anodine de la matrice de poids, ne signifie pas forcément que les corrélations entre chaque couple de stations soient correctement reproduites. C'est pour cette raison qu'il est nécessaire de choisir des matrices de poids décrivant les relations réalistes entre chaque couple de stations. L'utilisation des poids basés sur des fonctions de distance, par exemple, peut résulter en de faibles corrélations entre les stations concernées qui sont réellement fortement corrélées. Dans le cadre de cette thèse, des matrices intégrant les corrélations mensuelles entre les données de précipitations (article 1 au chapitre 3) et de températures (article 2 au chapitre 4) ont été utilisées et recommandées pour les études subséquentes.

Enfin, il paraît intéressant de souligner cet aspect « *de faire d'une pierre deux coups* » apporté par l'approche proposée. En effet, mettre l'emphasis uniquement sur l'autocorrélation spatiale a permis de reproduire aussi bien les autocorrélations spatiales observées sur

l'ensemble des stations que les corrélations entre chaque couple de stations. L'autocorrélation spatiale a été également utilisée pour régler le problème de l'intermittence spatiale des quantités de précipitations. Elle a été, aussi, un atout pour la génération des autres variables climatiques puisqu'elle a permis de reproduire la dépendance spatiale entre ces variables sans, toutefois, augmenter la dimension du modèle autorégressif (voir article 2 au chapitre 4). Le prochain chapitre présente l'application de cette approche de régionalisation pour la génération multi-site des données d'occurrences et de quantités de précipitations.

CHAPITRE 3

ARTICLE 1: STOCHASTIC MULTI-SITE GENERATION OF DAILY PRECIPITATION DATA USING SPATIAL AUTOCORRELATION

Abstract

There are a number of stochastic models which simulate weather data required for various water resources applications in hydrology, agriculture, ecosystem and climate change studies. However, many of them ignore the dependence between station locations exhibited by the observed meteorological time series.

This paper proposes a multi-site generation approach of daily precipitation data based on the concept of spatial autocorrelation. This theory refers to spatial dependence between observations with respect to their geographical adjacency. In hydrometeorology, spatial autocorrelation can be computed to describe daily dependence between the weather stations through the use of spatial weight matrix, which defines the degree of significance of the weather stations surrounding each observation.

The methodology is based on the use of the spatial moving average process to generate spatially autocorrelated random numbers which will be used in a stochastic weather generator. The resulting precipitation processes satisfy the daily spatial autocorrelations computed using the observed data. Monthly relationships between the spatial moving average coefficients and daily spatial autocorrelations of the precipitation processes have been developed to find the spatial moving average coefficients which reproduce the observed daily spatial autocorrelations in the synthetic precipitation processes.

To assess the effectiveness of the proposed methodology, seven stations in the Peribonca River Basin in the Canadian province of Quebec were used. The daily spatial autocorrelations of both precipitation occurrences and amounts were adequately reproduced, as well as the total monthly precipitations, the number of rainy days per month and the daily

precipitation variance. Using appropriate weight matrices, the proposed multi-site approach permits not only to reproduce the spatial autocorrelation of precipitation between the set of stations, but also the interstation correlation of precipitation between each pair of stations.

3.1 Introduction

Weather generators have been used successfully for a wide array of applications. They became increasingly used in various research topics, including more recently, climate change studies. They can generate series of climatic data with the same statistical properties as the observed ones. Furthermore, weather generators are able to produce series for any length of time. This allows developing various applications linked to extreme events, such as flood analyses.

Weather generators can also be used with downscaling approaches to create local scale climate scenarios from the global circulation models (GCMs) outputs. Indeed, the GCMs are limited for regional analysis by their low spatial resolution. Their results cannot be directly used on a local scale. For this reason, weather generators represent an attractive approach to generate precipitation and temperature time series on the watershed using the output of these models. Weather generators for assessing the effect of climate change can also generate series of climatic data based on climate change scenarios from GCMs.

Weather generators may be classified from a statistical point of view in two main categories. The first one is parametric such as the Richardson-type Weather Generator (WGEN) and the Long-Ashton Research Station Weather Generator (LARS-WG) (Semenov and Barrow 1997), developed respectively in the USA and Europe. WGEN generates precipitation occurrences using a first-order two-state Markov chain, while the precipitation amounts on rainy days are simulated by commonly used frequency distributions such as the gamma or exponential distributions. The precipitation occurrences are modelled in LARS-WG as alternate wet and dry series. The length of each sequence is chosen randomly from a semi-empirical distribution. For the wet days, the precipitation amounts are generated using other semi-empirical precipitation distributions.

The second type of weather generators uses non-parametric methods (Brandsma and Buishand 1997; Buishand and Brandsma 2001). This technique does not require any

assumption or parameter estimation and uses only the resampling of a weather variables vector on a day of interest from the historical data by conditioning on the simulated values of previous days. Resampling is done from the k nearest neighbours of these values using a weighted Euclidean distance.

Most of the existing weather generators are used at a single site independently of the others and thus ignoring the spatial dependence exhibited by the observed data. Some models, such as space-time models (Bardossy and Plate 1992; Bogardi *et al.* 1993) have been developed to regionalize the weather generators. In these models, the precipitation is linked to the atmospheric circulation patterns using conditional distributions and conditional spatial covariance functions. Another multi-site model called the non-homogeneous hidden Markov model (Bellone *et al.* 2000; Hughes and Guttorp 1994a, 1994b; Hughes *et al.* 1999) uses an unobserved weather state to link the large scale atmospheric measures and the small scale spatially discontinuous precipitation field. Although these two models preserve the relevant statistical information, they are complicated to implement and are unable to adequately reproduce the observed correlations. Buishand and Brandsma (2001) developed a multi-site approach based on the nearest-neighbour resampling. This approach does not deal with climate change, as the simulated series are based on the repetition of historical segments of data.

Wilks (1998) developed an approach based on serially independent but spatially correlated random numbers. For a network of k stations, a collection of $k(k-1)/2$ empirical curves must be developed for all possible pair stations, for each precipitation process and for each month. These curves link the random number correlations and the precipitation amount or occurrence correlations for every station pair and month. Even if this approach is able to take into account the spatial dependence of climatic data, it has the drawback of being computationally intensive. This approach also involves the matrix of interstation correlations, which is often not positive definite and computation is therefore not feasible. Moreover, for

large station networks this matrix becomes large and difficult to handle and all the coefficients of interstation correlations are not necessarily significant.

This article proposes an approach for multi-site generation of weather data based on the concept of spatial autocorrelation. The methodology was applied to the Péribonca River Basin and surrounding area, in the Canadian province of Quebec. The next section provides a definition of spatial autocorrelation. Section 3.3 describes the multi-site approach. The data set used for this study is presented in section 3.4 and the results obtained in the studied basin are presented and discussed in section 3.5.

3.2 Spatial autocorrelation definition

Spatial autocorrelation is a correlation among values of a single variable in reference to their geographical adjacency (Griffith 2003). In other words, it is a correlation of a variable with itself in a geographic space. Spatial autocorrelation is an important feature for geographical dependence. It is analogous to the serial autocorrelation approach, which indicates the strength of the correlation between a single variable with itself at different times. Serial autocorrelation is usually employed to simulate and forecast series exhibiting autoregressive characteristics.

Spatial dependence can arise because there are processes or phenomena that connect different locations and constrain the observations in a given place to be similar to those in a nearby area. Thus, the spatial autocorrelation determines the strength with which the occurrence of a process in a location affects other locations that are spatially contiguous. For example, the occurrence of a weather system may affect precipitations at several weather stations, and the opening of a freeway interchange may affect the prices of housing in nearby locations (Odland 1988).

The most interesting aspect of spatial autocorrelation is the weighing scheme, which highlights the neighbours influencing the location of interest and ignores or reduces the

influence of other locations that do not interact with it. Neighbouring geographical values around an observation may be defined by a spatial weight matrix. Literature provides several functions to express the geographic contiguity between locations (Odland 1988; Anselin and Bera, 1998; Anselin 1980; Murdoch *et al.* 1993).

It should be noted that contiguity between observations is not limited to the geographical distance but can also mean other information describing the relation between locations. According to Odland (1988), the flexibility in defining the weights makes spatial autocorrelation statistics a useful means of investigating alternative hypotheses about relations among places. Tobler (1970) mentioned in the first law of geography that «everything is related to everything else, but near things are more related than distant things». Thus, the spatial weights can represent distance metrics, as well as hypotheses about the spatial processes, which produce the data.

Much progress has been made principally in the last twenty years to come up with theoretical frameworks for spatial analysis. Generalized procedures have been developed to measure spatial autocorrelation (Cliff and Ord 1981; Hubert *et al.* 1981; Upton and Fingleton 1985). Various spatial statistics have been constructed to compute the spatial autocorrelation, such as Moran's I (Moran 1950; Odland 1988; Griffith 2003), presented below:

$$I = \frac{\sum_{i=1}^n (x_i - \bar{x}) \sum_{j=1}^n w_{ij} (x_j - \bar{x}) / \sum_{i=1}^n \sum_{j=1}^n w_{ij}}{\sum_{i=1}^n (x_i - \bar{x})^2 / n} \quad (3.1)$$

where x_i denotes the observed value at location i , \bar{x} is the average of the x_i over the n locations and w_{ij} is the spatial weight between two locations i and j . w_{ij} takes a non-zero value if the two locations are neighbours and zero otherwise. In the matrix form of the right-hand part of Eq. (3.1), all weights are stored in the spatial weight matrix. Usually, the

weights w_{ij} may be in a row-standardized form, which means that all weights in a row sum up to 1 and by convention $w_{ii} = 0$.

As defined by Moran's I , the spatial autocorrelation differs from the Pearson product moment correlation coefficient expression by using a single variable and the weight parameters. Therefore, the Pearson coefficient describes the relation between two given variables X and Y , while Moran's I indicates the dependence among values of a single variable taking into account their geographical locations. In this paper, the spatial autocorrelation will also be known as Moran's I or I .

Moran's I lies approximately between -1 and 1 (Griffith 2003). When the values are uncorrelated and arranged randomly over space, the Moran's I is equal to its expected value $-1/(n-1)$ (Moran 1950; Cliff and Ord 1981), n is the total number of locations. Moran's I larger than the expected value represents positive spatial autocorrelation, which means that geographically nearby values tend to be similar, and Moran's I smaller than the expected value represents negative spatial autocorrelation, which means that the values tend to be dissimilar.

An alternative statistic of spatial autocorrelation is Geary's C (Odland 1988; Griffith 2003), which is defined using another measure of covariation:

$$C = \frac{n-1}{2 \sum_i \sum_j w_{ij}} \frac{\sum_i \sum_j w_{ij} (x_i - x_j)^2}{\sum_i (x_i - \bar{x})^2} \quad (3.2)$$

The C values are between 0 and 2. The spatial autocorrelation is positive if C is lower than 1, negative if C is between 1 and 2 and null if C is equal to 1.

3.3 Methodology

3.3.1 Uni-site weather generator

A Richardson (1981)-type stochastic weather generator is used to simulate time series of weather data. In this uni-site weather generator, a first-order two-state Markov chain is used to simulate the daily precipitation occurrence $x_t(i)$ at site i on day t . A uniform $[0, 1]$ random number $u_t(i)$ is drawn and compared with a critical probability, which is equal to one of the transitional probabilities, depending on whether the previous day was wet or dry:

$$p_c(i) = \begin{cases} p_{01}(i), & \text{if } x_{t-1}(i) = 0 \\ p_{11}(i), & \text{if } x_{t-1}(i) = 1 \end{cases} \quad (3.3)$$

A wet day is simulated if the random number is smaller than this critical probability:

$$x_t(i) = \begin{cases} 1, & \text{if } u_t(i) \leq p_c(i) \\ 0, & \text{otherwise} \end{cases} \quad (3.4)$$

Another uniform $[0, 1]$ random number $v_t(i)$, independent from the previous one, is then used to simulate the synthetic precipitation amounts by inverting a probability distribution function of the precipitation amounts. The exponential distribution has frequently been used as a first approximation of the distribution of rainfall amounts (Todorovic and Woolhiser 1975) as well as the gamma distribution (Katz 1977). In the case of an exponential distribution, the cumulative density function can be defined as:

$$F[r(i)] = 1 - \exp(-\lambda(i)r(i)) \quad (3.5)$$

where $\lambda(i)$ is the parameter describing the exponential distribution function at site i . Therefore, the precipitation amounts $r_t(i)$ at site i can be computed as:

$$r_i(i) = -\ln(1 - v_i(i)) / \lambda(i) \quad (3.6)$$

3.3.2 Proposed multi-site generation approach

The proposed approach aims at using, in the weather generator, spatially autocorrelated random numbers whose spatial autocorrelations will reproduce the daily spatial autocorrelations in the synthetic precipitation processes similar to those of observations to which the weather generator has been fitted. For a given day, the single variable used in Moran's I (Eq. (3.1)) represents the values observed at the set of stations on that day. Note that the observed daily spatial autocorrelations to be reproduced are the averages of the daily spatial autocorrelations observed between the selected weather stations over their shared recording years. If all stations are wet or dry, the value 1 is assigned to the spatial autocorrelation of occurrence. This is also the case for the precipitation amounts if all stations are dry. The total of the random numbers for each of the occurrence and the amount time series is equal to the total number of stations used in the analysis.

3.3.2.1 Spatially autocorrelated random numbers model

The multi-site approach proposed in this paper deals with the spatial moving average process to obtain spatially autocorrelated random numbers. These random numbers will be used in the weather generator to simulate n spatially autocorrelated synthetic time series of occurrences and amounts at n stations. The spatial moving average process may be expressed as (Cliff and Ord 1981; Cressie 1993):

$$V = \gamma \times W \times u + u \quad (3.7)$$

Where $V(n,1)$ is a vector of n spatially autocorrelated random numbers to be used for n locations;

$W(n,n)$ is a weight matrix;

$u(n,1)$ is a vector of n independent and uniformly [0, 1] distributed random numbers;

γ is the moving average coefficient. The extreme eigenvalues of the weight matrix establish the range of this coefficient, which is $\left[\frac{-1}{w_{\max}}; \frac{-1}{w_{\min}} \right]$; where w_{\max} is the maximum positive eigenvalue and w_{\min} is the largest negative eigenvalue in absolute value.

The spatial moving average process contains a vector $u(n,1)$ of n independent and uniformly $[0, 1]$ distributed random numbers as well as the part which creates the spatial dependence between the $u(n,1)$ components. Hence, the spatial dependence of the resulting random numbers, components of the vector $V(n,1)$, depends on the γ value. In other words, for different values of this coefficient one can obtain vectors $V(n,1)$ of n random numbers exhibiting different degrees of spatial autocorrelation noted I_v computed by Moran indicator (Eq. (3.1)).

It is important to note that the random numbers resulting from the spatial moving average process are not necessarily uniformly distributed, nor necessarily lie between 0 and 1, whereas the weather generator requires the uniform $[0, 1]$ random numbers to be used in equations (3.4) and (3.6), as noted above. Therefore, a transformation is needed to turn these spatially autocorrelated numbers into uniformly $[0, 1]$ distributed ones preserving the required spatial autocorrelation. This transformation is carried out using the cumulative distribution function of the spatially autocorrelated numbers. In fact, for each value of γ , 1000 vectors $V(n,1)$ were generated and a histogram is plotted using all the spatially autocorrelated numbers. The cumulative frequency distribution is thus computed by using a cumulative sum of frequencies from the histogram. Such a sample size was considered to ensure that the result is affected as little as possible by the noise. The polynomial that fits the cumulative distribution function is used to find the cumulative probabilities of the spatially autocorrelated numbers. These values which occur between 0 and 1 will be used in the weather generator (Eqs. (3.4) and (3.6)) to model the occurrence and amount processes at each station.

Another issue that arises is what values of the coefficient γ should be used in equation (3.7) to obtain vectors of random numbers which will be fed in the weather generator to simulate the synthetic precipitation processes with daily spatial autocorrelations similar to those observed. Finding these values of the coefficient γ is the focus of the following analysis.

3.3.2.2 Moving average coefficients estimation

Using different values of the moving average coefficient γ , one can obtain vectors $V(n,1)$ enclosing n spatially autocorrelated random numbers with different spatial autocorrelations, indicated by I_V values, and thus the weather generator can model precipitation time series also displaying different spatial autocorrelations. Accordingly, the first step for the approach is to empirically establish the relationship between γ values and the I_V values of the resulting random numbers from the spatial moving average process. In fact, this operation consists in varying γ over its range $\left[\frac{-1}{w_{\max}}; \frac{-1}{w_{\min}} \right]$ and calculating Moran's I_V values of the resulting random numbers. Such a relationship is required for both occurrence and amount processes.

These random numbers generated from the spatial moving average process are then incorporated in the weather generator to simulate the precipitation occurrences at the stations. Each random number from the n components of $V(n,1)$ will be used for each of the n stations. This operation yields another curve linking the I_V values computed for the used random numbers and the I_X values of the resulting synthetic precipitation occurrences. I_X indicates the spatial autocorrelation computed by Moran's I for a vector X of n values of precipitation occurrence x generated at n stations:

$$X = [x(1) \quad \dots \quad x(n)] \quad (3.8)$$

In the case of the precipitation amounts, the spatial moving average process provides other random numbers whose spatial autocorrelations indicated by the I_V values are in relationship with the γ values. As was the case with the precipitation occurrences, a curve is also established between the I_V values of the used random numbers and the I_R values of the resulting synthetic precipitation amounts. I_R indicates the spatial autocorrelation computed by Moran's I for a vector R of n values of precipitation amount r generated at n stations:

$$R = [r(1) \quad \dots \quad r(n)] \quad (3.9)$$

It should be noted that the used γ values for the precipitation amount process will give a full vector $V(n,1)$ with non zero components from the spatial moving average process. Since each component of this vector will be used in the weather generator to produce the precipitation amount at each station, the simulated amounts will thus be different from zero. Therefore, the generated occurrence process will be used to change the precipitation amount at each dry station to zero. The spatial autocorrelations for the precipitation amounts as defined above are thus computed after this operation.

The next step is to find which values of coefficient γ should be used to generate synthetic time series with daily spatial autocorrelations similar to those of the observations for occurrence and amount processes. For each process, both relationships discussed above, namely (I_V, γ) and (I_X, I_V) or (I_R, I_V) will be used. One should identify on the one hand, the I_V values of the random numbers linked to the spatial autocorrelations observed for this process (I_X or I_R values), and on the other hand, the γ values linked to these I_V values. However, this procedure results in relationships between the γ values and spatial autocorrelations for each process, (I_X, γ) or (I_R, γ) , which can substitute for the two relationships mentioned above. Therefore, one can use these relationships directly to find the γ values which correspond to the observed values of I_X and I_R . Note that such

relationships can be established for each month of the year and the spatial autocorrelation values can be averaged over 1000 repetitions for high stability and low noise.

3.3.2.3 Data generation

Once γ values linked to the observed daily spatial autocorrelations of both precipitation occurrences and amounts are calculated from the monthly curves (I_X, γ) and (I_R, γ) , the generation phase starts. In fact, the spatial moving average model is reused with the selected γ values to give the spatially autocorrelated random numbers which will be inserted in the weather generator (Eqs. (3.4) and (3.6)) to produce synthetic meteorological time series exhibiting daily spatial autocorrelations similar to those of the observed ones. The length of the synthetic time series is specified by the user.

The flow chart in figure 3.1 summarizes this approach. It illustrates the parameter estimation and the generation phases for a given month. In this figure, m is the total number of γ values taken from its range and l is the total number of days in a given month.

3.4 Description of data set

The Péribonca River Basin (and surrounding area) (Figure 3.2) was used to test the approach. This watershed is located north of the Lac-Saint-Jean, in the Canadian province of Quebec, an area characterized by a wet climate, relatively cool summers and snow precipitations from November to April. The territory under study extends in latitude from 50°44' 00 "N to 48°37' 12"N and in longitude from 72°35' 35"W to 71°10' 06"W. Seven stations were selected for the analysis: Péribonca (P), Hémon (H), Bonnard (B), Chute-du-Diable (Cd), Chute-des-Passes (Cp), St-Léon-de-Labrecque (Sl) and Normandin CDA (N). The location of these stations is provided in Figure 3.2 and in Table 3.1. Table 3.2 gives the observed numbers of wet days per month per station and Table 3.3 shows the monthly average amounts of precipitation observed at each station. The data for each station are averaged over its recorded period.

The daily spatial autocorrelation values of occurrences and amounts depend on the series length employed. The user may opt for a 30-year duration, as recommended by the World Meteorological Organization (WMO). For the selected stations, such a period is not always available, with the recording durations varying from 17 to 57 years as indicated by Table 3.1. Nevertheless, the computation of the daily spatial autocorrelations was carried out using a shared recorded period for all stations, which is 14 years starting from 1963 to 1976. The weather generator parameters were calculated monthly at each station over its recorded period (Table 3.1) using the maximum likelihood estimator. The simulation period can be run for any length of time. The results presented are for a 50-year simulation period.

To realize the proposed multi-site approach, the weight matrix must be formalized. In this study, several matrices were tested using different techniques such as the Delaunay triangularization method (Delaunay 1934) and the inverse square distance method. The first method consists on the triangularization of the study area such that no station is inside the circumcircle of any triangle. Neighbours are defined as two points at vertexes of the same triangle. In the second method, the weight is inversely proportional to the square of the distance that separates each two locations. The resulting row-standardized weight matrix using the inverse square distance method is:

$$W = \begin{bmatrix} 0 & 0.0980 & 0.0053 & 0.4972 & 0.0154 & 0.1963 & 0.1879 \\ 0.1426 & 0 & 0.0087 & 0.0745 & 0.0226 & 0.0511 & 0.7006 \\ 0.0956 & 0.1082 & 0 & 0.0997 & 0.5103 & 0.0946 & 0.0916 \\ 0.2707 & 0.0279 & 0.0030 & 0 & 0.0094 & 0.6494 & 0.0396 \\ 0.1462 & 0.1475 & 0.2677 & 0.1646 & 0 & 0.1523 & 0.1217 \\ 0.1314 & 0.0235 & 0.0035 & 0.7986 & 0.0107 & 0 & 0.0321 \\ 0.2325 & 0.5960 & 0.0063 & 0.0901 & 0.0158 & 0.0594 & 0 \end{bmatrix} \quad (3.10)$$

Each element w_{ij} indicates the weight of the pair stations (i, j) and by convention $w_{ii} = 0$. The stations used are classified as mentioned above. Thus, the non-zero elements of the first row represent the weights between the Péribonca location and the remaining stations Hémon, Bonnard, Chute du Diable, Chute des passes, St-Léon-de-Labrecque and Normandin CDA,

respectively. The second row contains the weights between the Hémon station and the other stations according to the order noted above and so on. It appears that the weights of the pair stations including Bonnard or Chute-des-passes stations stored in the third and the fifth columns, respectively, are low compared to the others, which means that Bonnard and Chute-des-passes, which are actually located far from the remaining stations, are not deemed their nearest neighbours according to the inverse square distance method. The third and the fifth rows related to Bonnard and Chute-des-passes stations respectively indicate that the pairs of these stations, which are the closest, have the most important weights.

Although these techniques based on geographical information perform well in representing the arrangement of the weather stations in the watershed, it is more general to select the spatial weights which describe the accommodating hypotheses about the relations among weather stations as mentioned by (Odland 1988). In fact, according to the tests carried out for this study, the low weight values assigned to some pairs of stations by the methods mentioned above result in a loss of correlation (Pearson product moment correlation) observed between the stations which compose those pairs (example: the correlations between Bonnard station and the remaining stations excluding Chute-des-passes).

The mechanical definition of spatial weights using these methods based on the distance criterion is thus not appropriate in this case because all stations are significantly correlated. Therefore, it seems more realistic to select the spatial weights, which describe the dependence between the station pairs. The proper weight matrix should allow the reproduction of not only the spatial autocorrelations over the entire watershed, but also the standard correlation between every station pair. For this reason, the weight matrix is formalized using the annual correlations of precipitation between each pair of stations. Note that these correlations are those of Pearson. Furthermore, the weight matrix is row-standardized. The resulting row-standardized weight matrix using the annual correlations of precipitation between each pair of stations is:

$$W = \begin{bmatrix} 0 & 0.1661 & 0.1119 & 0.2009 & 0.1360 & 0.1915 & 0.1936 \\ 0.1836 & 0 & 0.1226 & 0.1712 & 0.1482 & 0.1630 & 0.2114 \\ 0.1556 & 0.1542 & 0 & 0.1672 & 0.2088 & 0.1596 & 0.1546 \\ 0.1983 & 0.1529 & 0.1187 & 0 & 0.1476 & 0.2051 & 0.1775 \\ 0.1621 & 0.1598 & 0.1790 & 0.1782 & 0 & 0.1565 & 0.1643 \\ 0.1987 & 0.1530 & 0.1191 & 0.2155 & 0.1362 & 0 & 0.1775 \\ 0.1967 & 0.1942 & 0.1129 & 0.1826 & 0.1400 & 0.1737 & 0 \end{bmatrix} \quad (3.11)$$

3.5 Results and discussion

As mentioned above, the extreme eigenvalues of the weight matrix establish the range of the moving average coefficient γ , which is $]-1; 4.6795[$ for the used weight matrix. The coefficients γ that are out of this range may not be in perfect relationship with the spatial autocorrelations of the resulting precipitation processes. Consequently, the largest daily spatial autocorrelations may not be adequately reproduced. This problem can be solved using the iterated spatial moving average process to increase the spatial autocorrelations of the resulting random numbers, as well as those of the synthetic precipitation processes. One can use two or more successive moving average processes.

Operationally, the analysis begins with modelling the precipitation occurrences. Precipitation amounts are then modeled given the rainy days. For occurrence process, the relationship between the γ values and I_x values of the occurrence series are established and represented adequately by a 3-order polynomial function such as:

$$\gamma = a I_x^3 + b I_x^2 + c I_x + d \quad (3.12)$$

where a , b , c and d are empirical parameters that describe the relationship, determined by a regression fit for each month. The 3-order polynomial function has been chosen because it produces the best fit to a data set. Such a curve is established for each month of the year. Figure 3.3 gives this relationship for June. Similar curves were obtained for the other months. The curve presented in Figure 3.3 can be adequately represented by this equation:

$$\gamma = 11.972 I_X^3 - 10.777 I_X^2 + 5.8183 I_X + 0.9937 \quad (3.13)$$

For example to reproduce an observed spatial autocorrelation of 0.6, the required γ value to be used in the moving average process is 3.19.

The γ values linked to all observed values of I_X for each month may be extracted from these curves to generate the spatially autocorrelated random numbers from the moving average process. These random numbers will be used in the weather generator to produce synthetic precipitation occurrences with daily spatial autocorrelations similar to those observed between the seven stations.

Figure 3.4 illustrates the observed versus simulated monthly numbers of rainy days at all stations for January, April, July and October. The observed data are averaged over the recorded period for each station and the simulated ones are averaged over the simulated period which is 50 years. Similar results are obtained for the other months. Clearly, the monthly numbers of rainy days are adequately reproduced. The multi-site approach succeeded in modeling the occurrence processes using the weather generator at the seven stations and for all months.

Figures 3.5a and 3.5b illustrate the joint probabilities of the dry and wet states respectively for each pair of stations and each month. It appears that the joint probabilities of the occurrence process are also well reproduced by the multi-site generator ($R^2=0.8936$ for the dry state and $R^2=0.8681$ for the wet state).

The performance of the multi-site approach in reproducing the observed spatial dependence between the seven stations on each day was also investigated. Figure 3.6 shows the observed and simulated daily spatial autocorrelations of precipitation occurrences. The observed ones are averaged over the shared recorded period between the seven stations and the simulated ones are averaged over the simulation period. Therefore, 365 points representing the coordinate pairs are plotted in this graph. This figure indicates that the spatial

autocorrelations, exhibited by the seven stations on each day, are adequately reproduced by the multi-site approach ($R^2=0.8508$).

Given the wet days, precipitation amounts can be handled in a similar fashion as for the occurrences. The relationship between the γ values and I_R values of precipitation amounts are also represented by a 3-order polynomial function such as:

$$\gamma = e I_R^3 + f I_R^2 + g I_R + h \quad (3.14)$$

where e, f, g and h are parameters provided by a regression fit for each month. Such a curve will also be established for each month. Figure 3.7 shows such a relationship for January. Similar curves were established for the other months.

The γ values linked to all observed values of I_R for each month can be found from these curves to generate the spatially autocorrelated random numbers from the moving average process. These random numbers will be used in the weather generator to produce synthetic precipitation amounts on the wet days with daily spatial autocorrelations similar to those observed.

Figure 3.8 illustrates the total monthly precipitations at the seven stations for February, June, September and December. The observed and simulated total monthly precipitations are generally in good agreement. A similar result is obtained for the other months. Figure 3.9 provides a scatter plot of observed and simulated daily spatial autocorrelations of precipitation amounts. It appears that the daily spatial autocorrelations of precipitation amounts are satisfactorily reproduced over the year ($R^2=0.757$). However, the daily spatial autocorrelations of occurrence process are better reproduced than those of amount process because the generation of amounts is influenced by the spatial autocorrelations of occurrence process generated before.

The multi-site approach adequately reproduces the daily spatial dependence of both precipitation occurrences and amounts, as well as the total monthly precipitations and the monthly numbers of rainy days. However, it would be interesting to investigate how this approach reproduces the precipitation variation. Figure 3.10 illustrates the observed and simulated daily precipitation standard deviations for January, April, August and October. The standard deviations are computed for all days of each month over the 50-year simulation period. This figure shows that the multi-site approach successfully reproduces the standard deviations. Similar results are obtained for the remaining months.

It will also be interesting to investigate if the approach reproduces the Pearson correlations of occurrences and amounts between every pair of stations. Figure 3.11 shows the interstation correlations between daily precipitation occurrences for all pairs of stations and months. The observed ones are calculated over the shared recorded period between each pair of stations and the simulated ones are calculated over the simulation period. In this figure, it seems that some of the larger interstation correlations are underestimated. Figure 3.12 illustrates the interstation correlations between daily precipitation amounts for all pairs of stations and months. In this case, the interstation correlations seem to be overestimated. This behavior of the synthetic interstation correlations is expected because the multi-site approach did not try to reproduce exactly the observed interstation correlations despite the good result obtained for the occurrences ($R^2=0.5555$ for the occurrences and $R^2=0.1259$ for the amounts).

To improve this result, monthly weight matrices representing the monthly interstation correlations of precipitation between all pairs of stations are used for each month. Taking the spatial weights to the 2th power for the occurrences and the 3th power for the amounts was necessary to optimize the results. Figures 3.13 and 3.14 show these results. It can be seen that the monthly weight matrices improve the estimation of the interstation correlations for the precipitation processes, thus highlighting the importance of selecting the proper weight matrix ($R^2=0.7989$ for the occurrences and $R^2=0.791$ for the amounts). The remaining aberrations can be attributed to the sample size, which is 14 years used to calibrate the

proposed approach. Note that these interstation correlations have been reproduced automatically in the synthetic time series by reproducing the daily spatial autocorrelations.

The spatial moving average process used for this multi-site approach consists on simple computations that do not require spatial weight matrix operations. Therefore, the proposed approach does not present any disadvantage for a large data set. Moreover, the spatial weight matrix tends to be sparse, which facilitates the computations, and does not need a lot of memory. From a geographic point of view, for a large data set, the spatial autocorrelation becomes more interesting to investigate because the spatial patterns are more likely to develop over an extended region. Using the proposed approach, the Moran's I summarizes the complete spatial dependence into a single number, and only 12 relationships have to be developed to carry out the multi-site generation for each of the occurrence and amount process for all months. This number does not increase if the number of stations increases.

Because the multi-site approach proposed in this paper uses the spatial autocorrelation concept, which summarizes the complete spatial dependence among the set of stations into a single number, it appears interesting to compare its results to those obtained with spatially correlated random numbers used in Wilks (1998). In this second case, a number of $k(k-1)/2$ curves have to be developed for a network of k stations to reproduce the interstation correlations observed between each pair of stations in each month.

For the Peribonca River Basin, 504 empirical curves have been developed for all pairs of stations, for both precipitation occurrences and amounts and for all months. Figure 3.15 shows an example of such empirical curve obtained for the Peribonca and Chute du Diable stations in January to find which correlation of random numbers must be used to reproduce the precipitation occurrence correlation observed between these two stations in this month.

Figures 3.16 and 3.17 show the interstation correlations of daily precipitation occurrences and amounts, respectively, for all pairs of stations and months obtained using the spatially correlated random numbers. Since the aim of the 504 empirical curves is to reproduce the

precipitation correlations observed between each pair of stations in each month, the result is relatively better than that of the proposed approach obtained automatically via the spatial autocorrelation concept ($R^2=0.8948$ vs 0.7989 for the occurrences and $R^2=0.8651$ vs 0.791 for the amounts). However, comparing the observed versus simulated monthly spatial autocorrelations, Figures 3.18 and 3.19 indicate that the spatial autocorrelations observed in the study area in each month are well reproduced by the proposed approach especially for the occurrences ($R^2=0.9701$ vs 0.6302 for the occurrences and $R^2=0.8455$ vs 0.7461 for the amounts). Note that the recorded periods used to calibrate the two approaches are different. Indeed, the proposed approach uses the shared recorded period between the set of stations which is 14 years while the spatially correlated random number approach uses the shared recorded periods between each pair of stations whose minimum is 14 years.

Furthermore, the daily spatial autocorrelations are not reproduced by the spatially correlated random number approach because it works monthly to reproduce the interstation correlations. For the other statistical criteria presented above, the two approaches give practically similar results.

3.6 Conclusion

A multi-site generation approach of daily precipitation data was presented. It is based on the notion of spatial autocorrelation, which is the correlation between values of a single variable through space. The aim of this approach is to introduce in a weather generator, series of random numbers with spatial autocorrelations that can reproduce the observed daily spatial autocorrelations of precipitation processes, namely occurrences and amounts.

The important aspect in the definition of the spatial autocorrelation is the weight matrix, which gives a score to identify neighbouring observations. It should be noted that the values of Moran's indicator depend on the weighting scheme. Thus, the observed daily spatial autocorrelations may be changed using alternative neighbourhood criteria. This suppleness in establishing the weight matrix allows for the investigation of different hypotheses about the

spatial pattern and the differences in the obtained results can be a measure of the advantages of one hypothesis over the other.

The Péribonca River Basin and surrounding area in the Canadian province of Quebec were used to test this regionalization approach. Relationships were found between the spatial moving average coefficients, used to generate spatially autocorrelated random numbers, and the spatial autocorrelations for the precipitation processes computed by the Moran's indicator. These relationships provide the moving average coefficients which enable the observed daily spatial autocorrelations to be reproduced in the synthetic precipitation processes.

Results indicate that this approach is successful in reproducing the daily and monthly spatial autocorrelations between the seven stations used in the Péribonca watershed. A good agreement was also found between the observed and simulated monthly numbers of rainy days, the total monthly precipitation amounts as well as the daily precipitation variance.

It is important to emphasize that using the appropriate weight matrix representing the degree of significance of the weather stations surrounding each observation allows the multi-site approach to reproduce not only the daily spatial autocorrelations over the watershed, but also the monthly interstation correlations of precipitation processes between each pair of stations in each month.

A comparison has been done between the proposed approach using spatially autocorrelated random numbers and the approach of spatially correlated random numbers developed by Wilks (1998). While this latter leads to relatively more precision in reproducing the monthly interstation correlations, the monthly spatial autocorrelations are well reproduced by the proposed approach.

Finally, the spatial autocorrelation is a useful tool to describe spatial dependence of data in a geographic space. Using spatial autocorrelation for multi-site generation of precipitation

processes is successful. This approach has the advantage of being versatile and easy to implement, requiring only 12 empirical curves linking the moving average coefficient values to the spatial autocorrelation values of each of the precipitation occurrences and amounts and does not apply any restrictions on the spatial weight matrix. For a large data set, the proposed approach remains easy to execute because the overall pattern in the data is summarized in a single statistic. Moreover, the spatial moving average process used in this paper consists on a simple computation process.

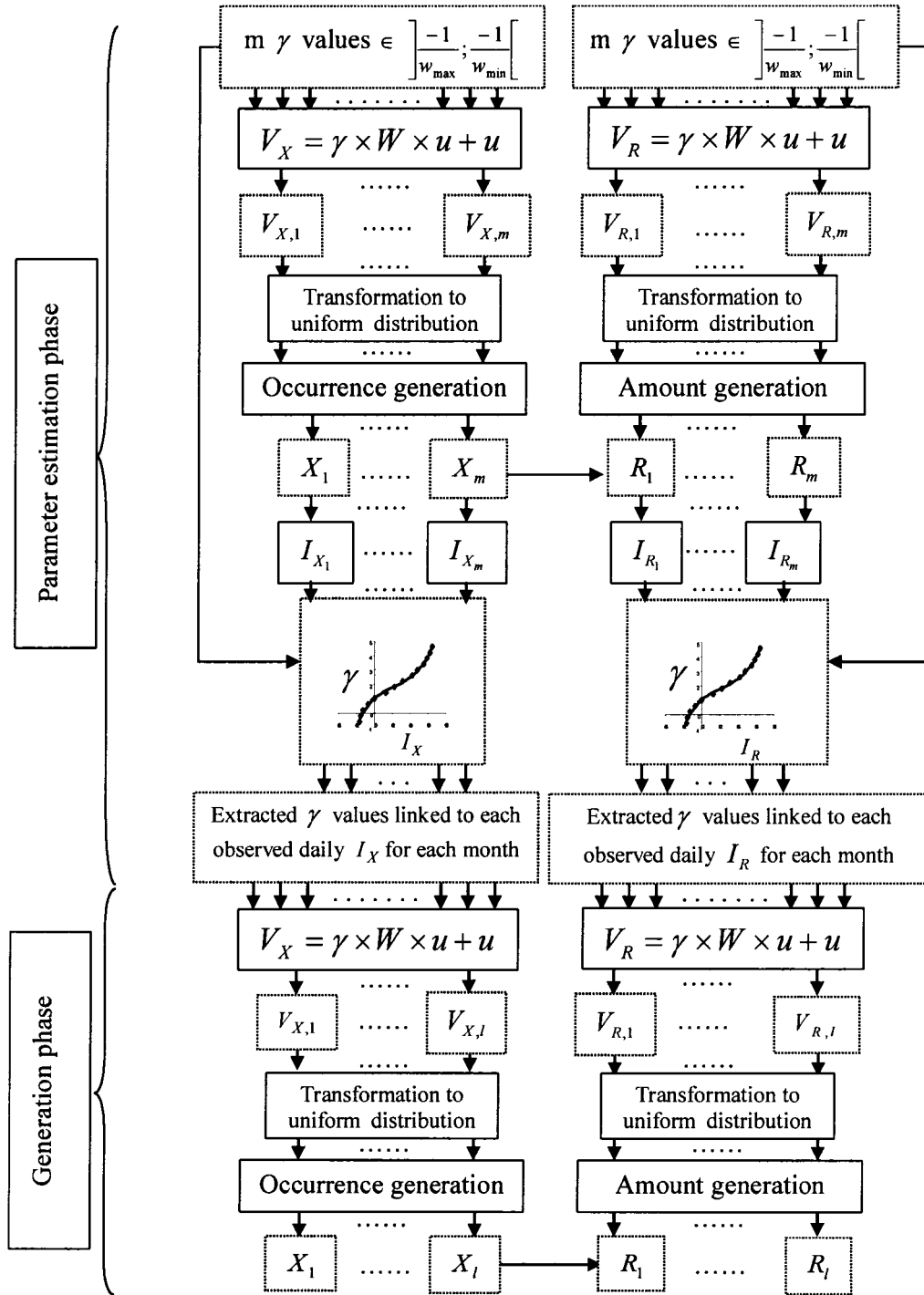


Figure 3.1 Flow chart for the proposed multi-site generation approach.

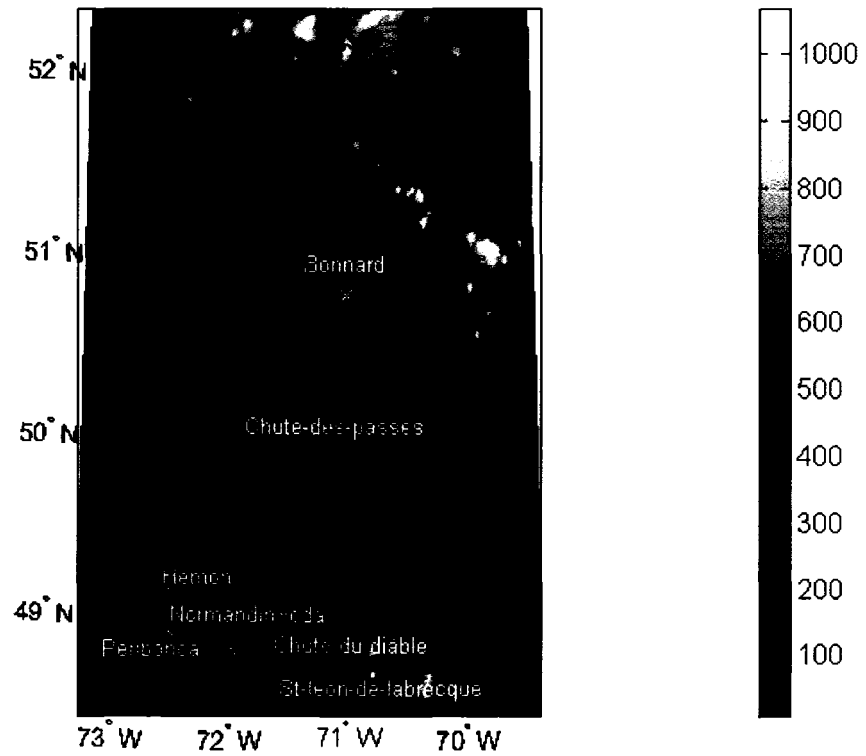


Figure 3.2 *The Peribonca watershed, showing the location of meteorological stations. Elevations in meters.*

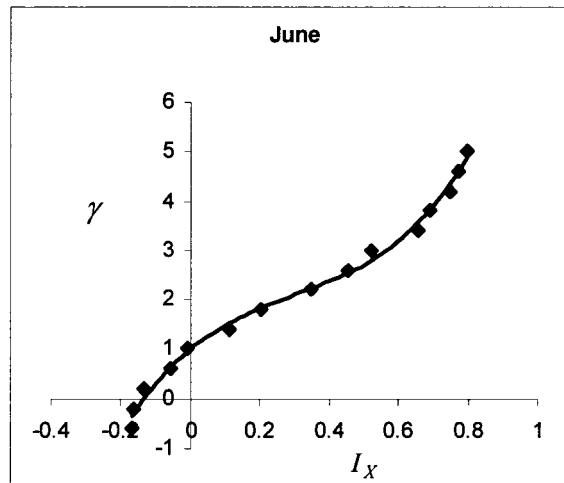


Figure 3.3 *Relationship between the moving average coefficients γ and Moran's I_X computed for the precipitation occurrences at the seven stations for June.*

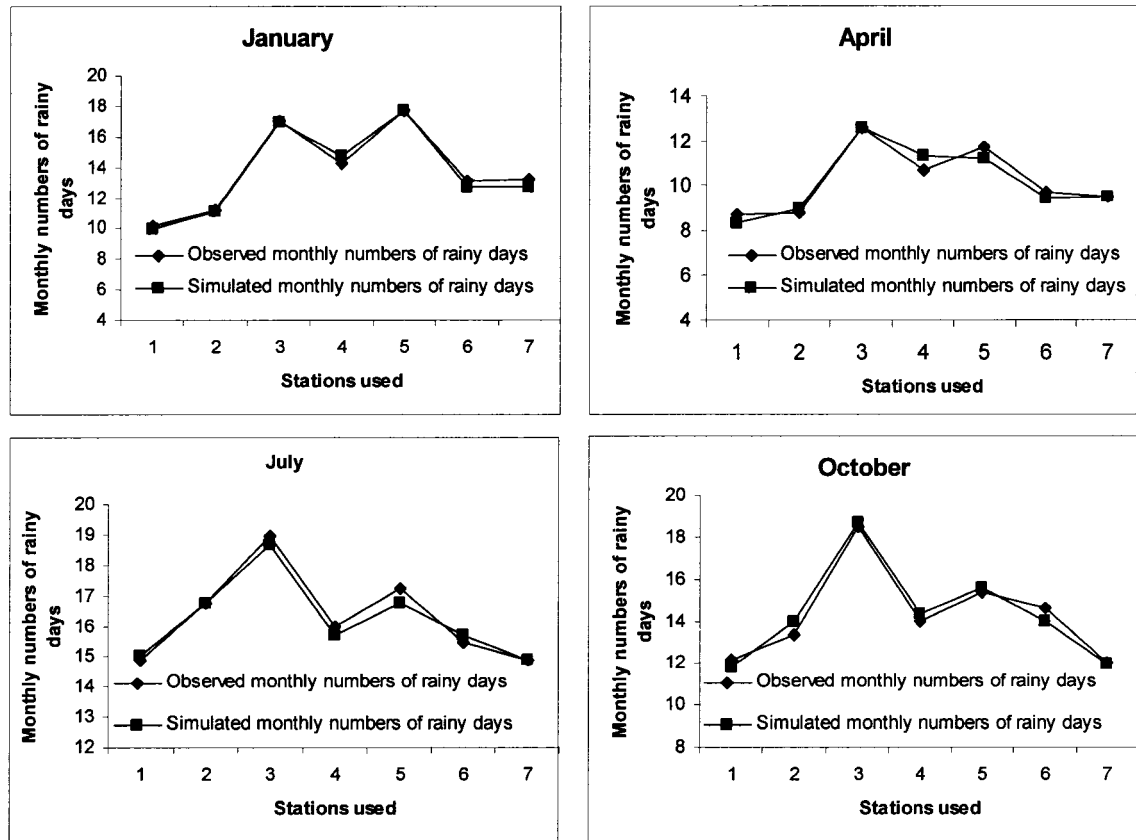


Figure 3.4 *Observed versus simulated monthly numbers of rainy days at all stations for January, April, July and October.*

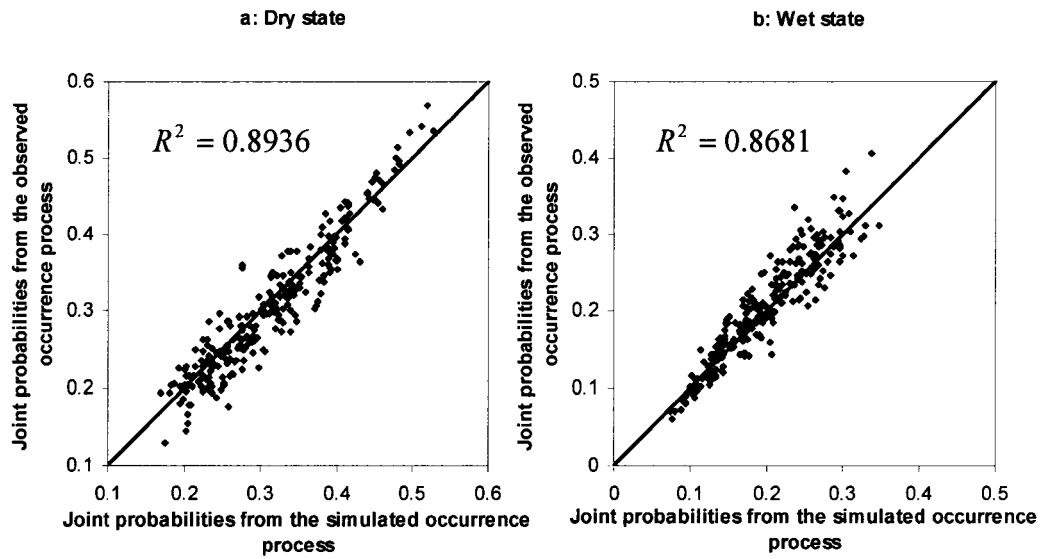


Figure 3.5 *Joint probabilities from observed and simulated occurrence process for each pair of stations and each month. Figure 5a for both stations are dry and figure 5b for both stations are wet.*

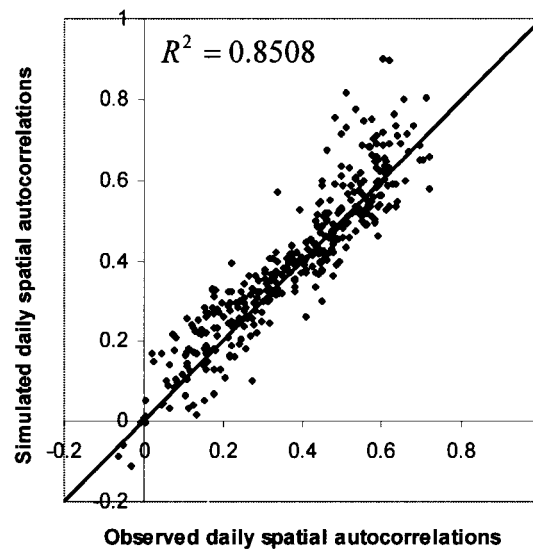


Figure 3.6 *Observed and simulated daily spatial autocorrelations I_x of precipitation occurrences.*

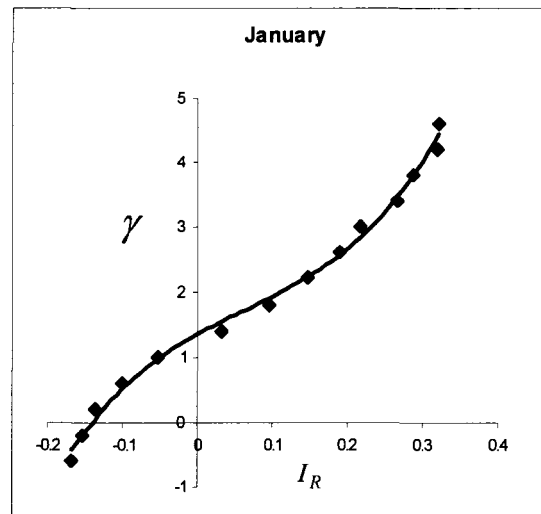


Figure 3.7 *Relationship between the moving average coefficients γ and Moran's I_R computed for the precipitation amounts at the seven stations for January.*

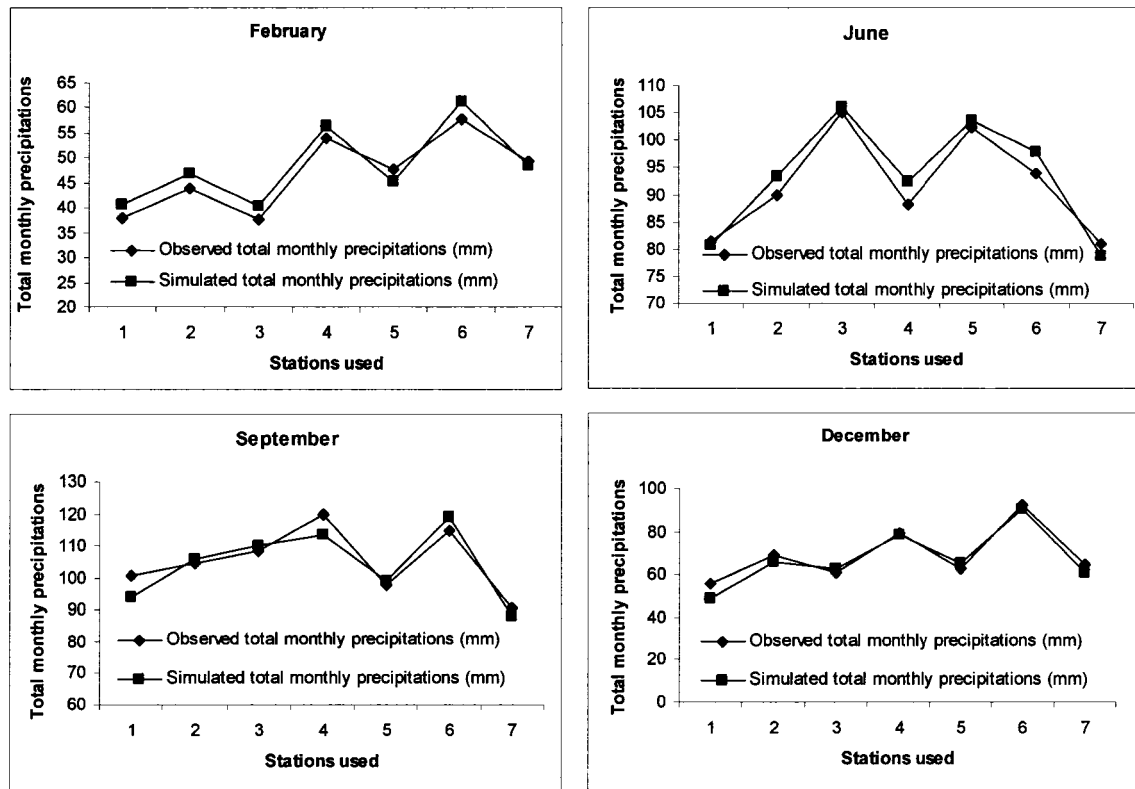


Figure 3.8 *Observed and simulated total monthly precipitations at the seven stations for February, June, September and December.*

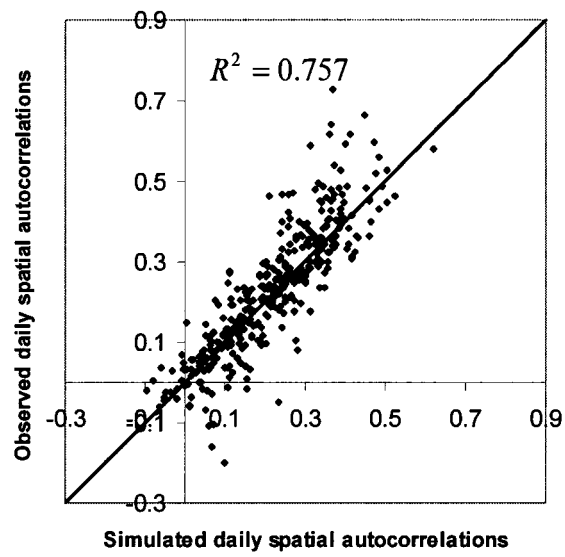


Figure 3.9 *Observed and simulated daily spatial autocorrelations I_R of precipitation amounts.*

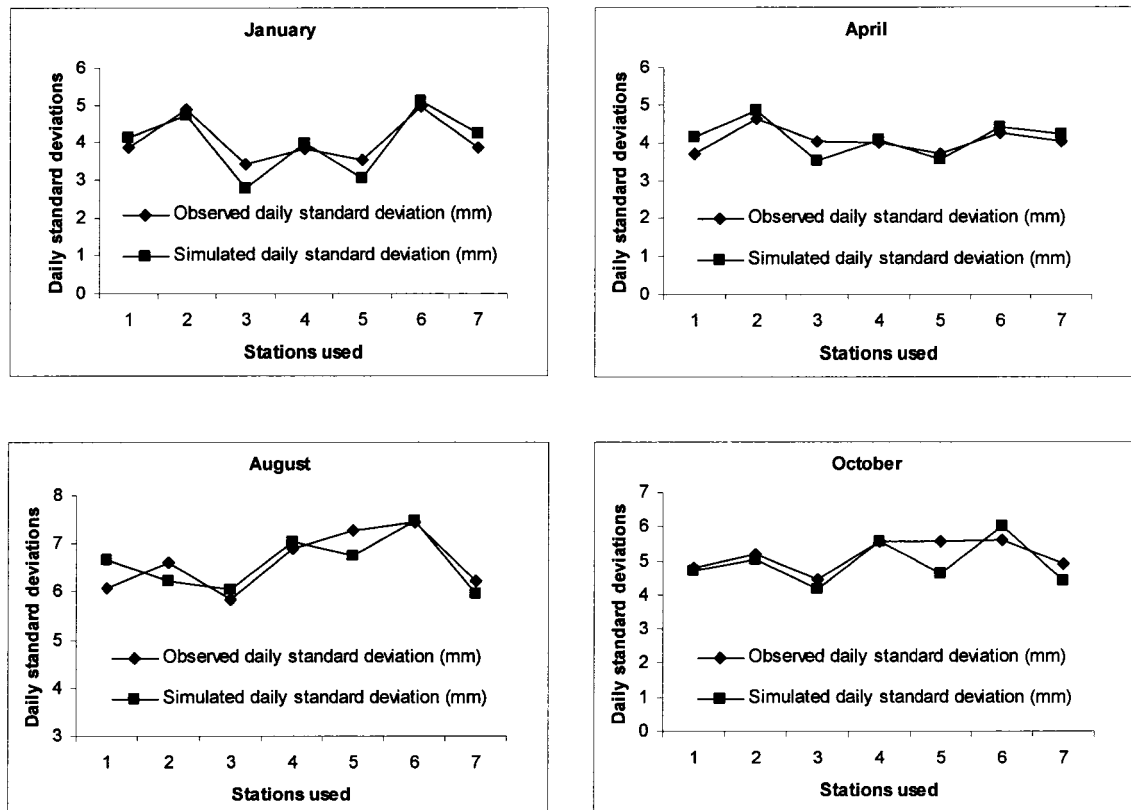


Figure 3.10 *Observed and simulated daily precipitation standard deviations at the seven stations for January, April, August and October.*

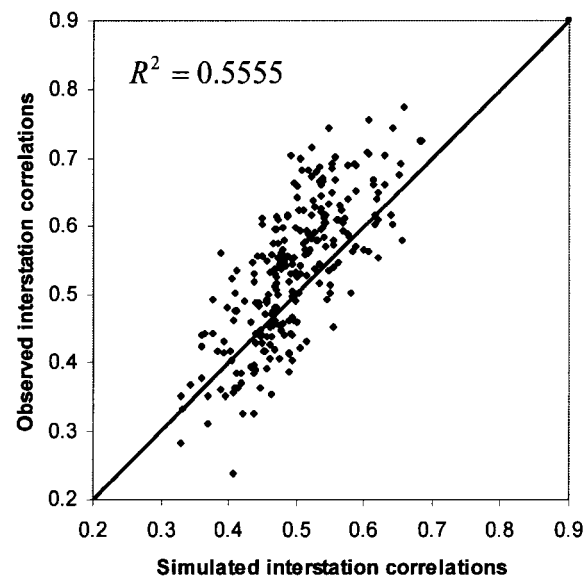


Figure 3.11 *Observed and simulated interstation correlations between daily precipitation occurrences for all pairs of stations and months.*

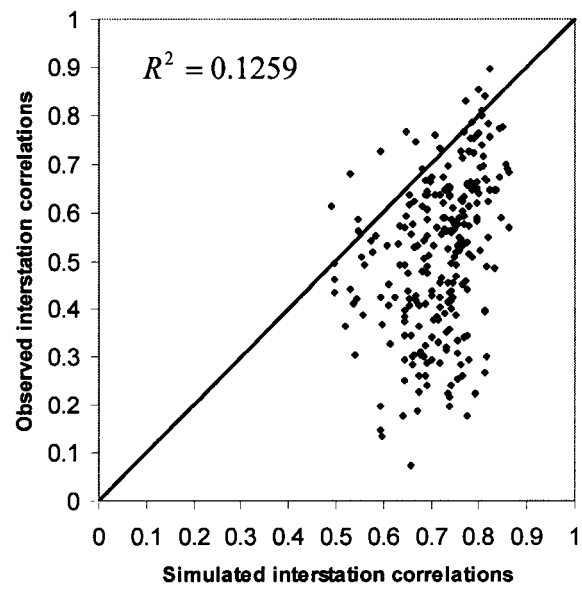


Figure 3.12 *Observed and simulated interstation correlations between daily precipitation amounts for all pairs of stations and months.*

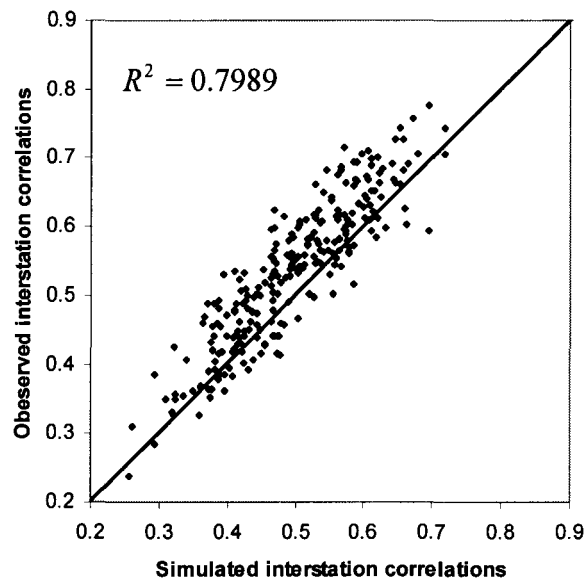


Figure 3.13 *Observed and simulated interstation correlations between daily precipitation occurrences for all pairs of stations and months using monthly weight matrices.*

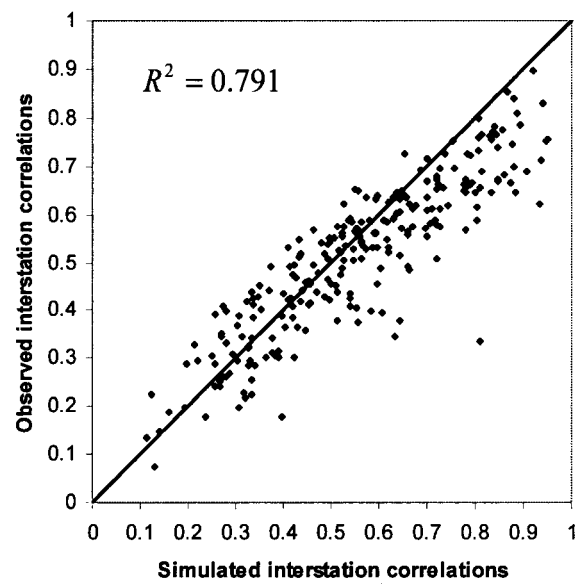


Figure 3.14 *Observed and simulated interstation correlations between daily precipitation amounts for all pairs of stations and months using monthly weight matrices.*

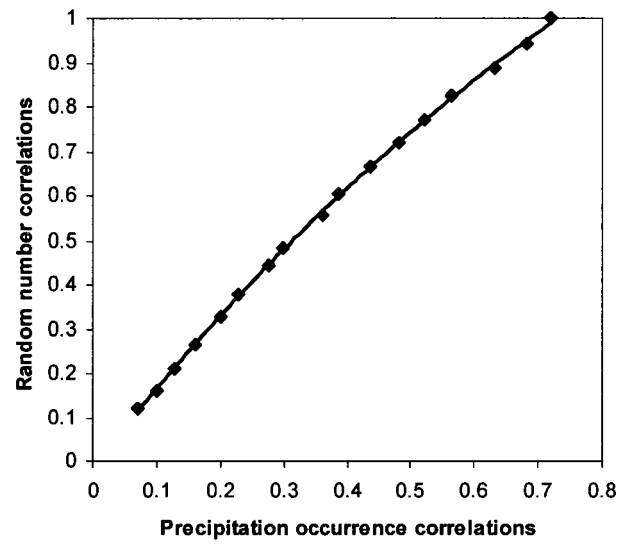


Figure 3.15 *Relationship between precipitation occurrence correlations and random number correlations for Péribonca and Chute du Diable stations in January.*

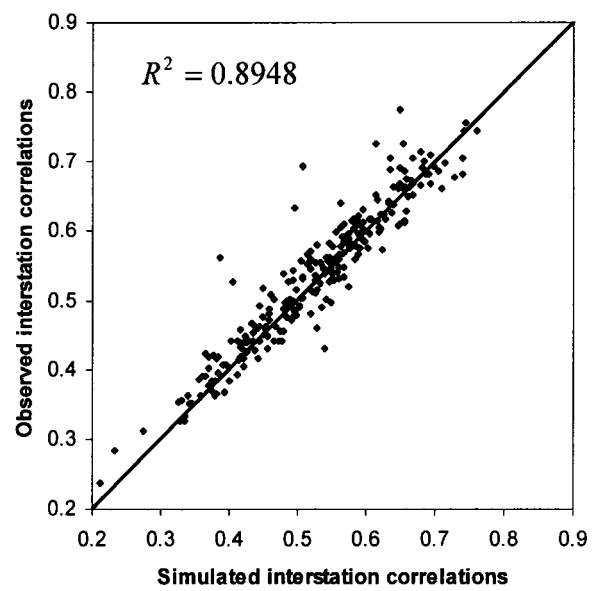


Figure 3.16 *Observed and simulated interstation correlations between daily precipitation occurrences for all pairs of stations and months using the Wilks' approach.*

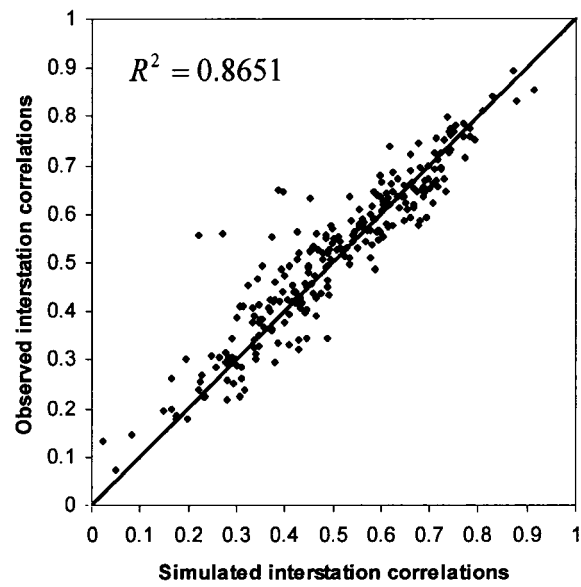


Figure 3.17 *Observed and simulated interstation correlations between daily precipitation amounts for all pairs of stations and months using the Wilks' approach.*

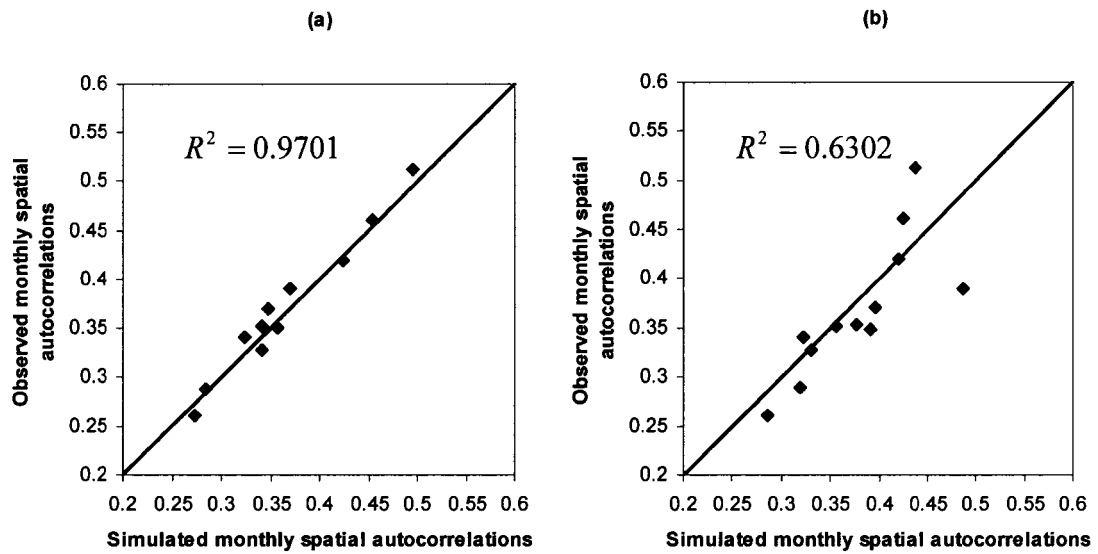


Figure 3.18 *Observed and simulated monthly spatial autocorrelations of precipitation occurrences using (a) the proposed approach and (b) the Wilks' approach.*

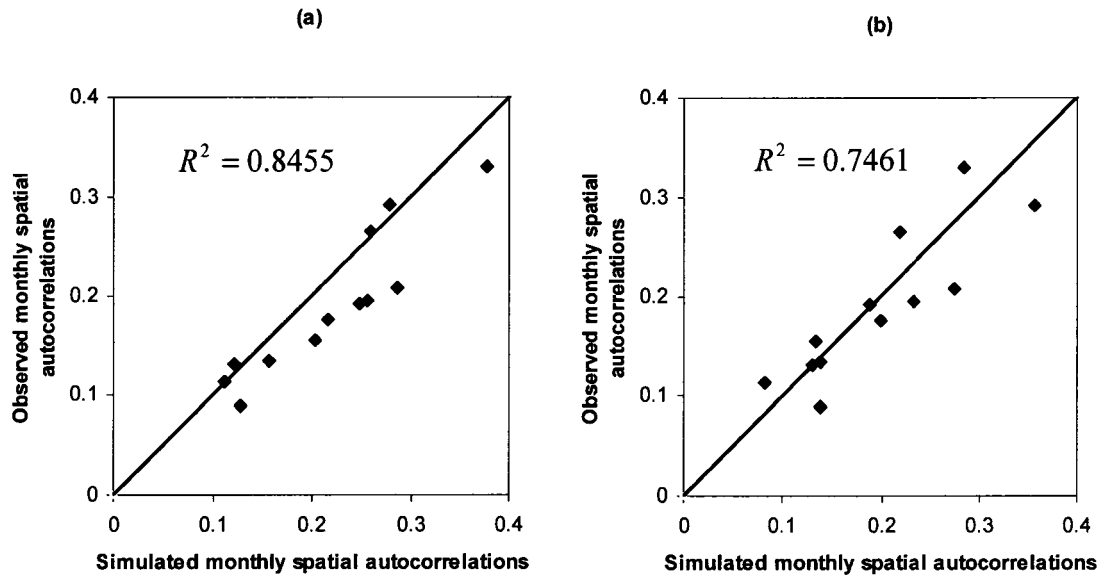


Figure 3.19 *Observed and simulated monthly spatial autocorrelations of precipitation amounts using (a) the proposed approach and (b) the Wilks' approach.*

Tableau 3.1

Location and recorded years of the used stations

Stations	Lat	Long	Elevation (m)	Recorded years
Péribonca	48°45'42"N	72°01'24"W	107	1951-2002
Hémon	49°03'48"N	72°35'35"W	206	1963-2002
Chute-du-Diable	48°45'00"N	71°42'00"W	174	1951-1976
St-Léon de Labrecque	48°40'26"N	71°31'33"W	131	1963-1997
Normandin CDA	48°51'00"N	72°32'00"W	137	1936-1992
Chute-des-Passes	49°50'25"N	71°10'06"W	398	1960-1976
Bonnard	50°44'00"N	71°02'00"W	506	1961-2000

Tableau 3.2

Observed numbers of wet days per month per station

Numbers of wet days												
Stations	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sept	Oct	Nov	Dec
P	10	8	8	8	12	12	15	14	14	12	11	11
H	12	8	8	9	13	14	17	16	16	13	11	11
Cd	14	12	12	11	13	13	16	15	15	14	15	15
Sl	13	10	9	10	12	13	15	15	15	15	13	14
N	13	10	10	9	12	13	15	14	13	12	12	13
Cp	18	15	13	12	13	13	17	17	16	15	18	16
B	17	13	13	13	14	16	19	18	18	18	19	18

Tableau 3.3

Monthly average amounts of precipitation observed at each station

Monthly average precipitation amounts per station (mm)												
Stations	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sept	Oct	Nov	Dec
P	54.6	38.1	43.1	46.5	73.5	81.3	113.2	96.1	101	71	58	56
H	67.6	44	48.7	55.8	86.5	90	123.1	105.7	105	81	73	69
Cd	64.5	53.8	60.5	49.8	76.9	87.9	120.6	108.5	120	84	85	79.1
SI	75	57.7	55.5	58.8	88	93.9	137	118.1	115	100	91	92.4
N	57	49.3	53	50.7	71.3	81	108	88	91	63	60	64.8
Cp	58	47.6	49.3	48	73.5	102.3	142	119	99	84	88	62.3
B	54.3	37.8	45.9	56.1	75.6	105	136.1	106	108.3	77	78	60.5

CHAPITRE 4

ARTICLE 2: STOCHASTIC MULTI-SITE GENERATION OF DAILY WEATHER DATA

Abstract

Spatial autocorrelation is a correlation between the values of a single variable, considering their geographical locations. This concept has been used for the multi-site generation of daily precipitation data. An extension of this approach is presented in this paper. It aims firstly to obtain an accurate reproduction of the spatial intermittence property in the synthetic precipitation amounts, and then to extend the multi-site approach to the generation of other daily weather data, particularly maximum and minimum temperatures.

Monthly spatial exponential functions have been developed for each weather station according to the spatial dependence of the occurrence processes over the watershed to fulfill the spatial intermittence condition in the synthetic time series of precipitation amounts. As was the case for the precipitation processes, the multi-site generation of daily maximum and minimum temperature data was realized using spatially autocorrelated random numbers. These random numbers are incorporated in the weakly stationary generating process used in the Richardson weather generator, with no modifications made. Suitable spatial autocorrelations of the random numbers allow the reproduction of the observed daily spatial autocorrelations and monthly correlations between each pair of stations.

Peribonca River Basin watershed is used to test the performance of the proposed approaches. Results indicate that the spatial exponential functions succeeded to reproduce an accurate spatial intermittence in the synthetic precipitation amounts. The multi-site generation approach was successfully applied for temperature data. These weather data were adequately generated maintaining efficient daily spatial autocorrelations and monthly interstation correlations.

4.1 Introduction

Weather generators are increasingly being used in a variety of water resource studies due to their ability to provide series of weather data for any length of time with similar statistics as those to which the weather generators have been fitted. One of the most important applications concerns climate change. Weather generator parameters can be perturbed to compute future time series based on the climate change scenarios derived from GCMs (Global Circulation Models). The long synthetic time series can also allow an accurate analysis of the extreme events sought for many hydrological studies.

The most commonly used weather generator is that of Richardson (Richardson 1981; Richardson and Wright 1984) using a first-order two-state Markov chain model to generate precipitation occurrences, and a distribution function such as gamma or exponential to model the precipitation amounts. Another basic type of weather generator is the one of Semenov and Barrow (1997), which uses semi-empirical distributions to simulate the precipitation processes. There are also non-parametric weather generators (Brandsma and Buishand 1997; Buishand and Brandsma 2001) using the resampling of a weather variables vector on a day of interest from the historical data by conditioning on the simulated values of previous days. Wilks and Wilby (1999) present a good review of weather generators.

Each of the weather generators cited above operates at a single site, and thus fails to reproduce the spatial dependence exhibited by the observed data. In fact, the extension of the weather event over the watershed implies that the weather data at a given weather station is unlikely to be not correlated with those in the surrounding area. Further researches have therefore been conducted to develop multi-site weather generators such as space-time models (Bardossy and Plate 1992; Bogardi *et al.* 1993), which use the atmospheric circulation patterns with conditional distributions and conditional spatial covariance functions. Non-homogeneous hidden Markov model (Bellone *et al.* 2000; Hughes and Guttorp 1994a, 1994b; Hughes *et al.* 1999) uses a discrete weather state to link the large scale atmospheric measures and the small scale spatially discontinuous precipitation field. A multi-site weather generator

based on the nearest-neighbour resampling from the historical data has been developed by Buishand and Brandsma (2001). Wilks (1998) developed a multi-site version of the Richardson weather generator based on serially independent but spatially correlated random numbers. Further studies have been conducted to extend these multi-site approaches for temperature and solar radiation (Wilks 1999; Buishand and Brandsma 2001).

The multi-site weather generators presented above have different strengths and weaknesses. For instance, the space-time models and the non-homogeneous hidden Markov model are designed using relevant statistic information, but yield difficulties in their implementation with certain result limitations. The nearest-neighbour resampling model uses the historical record to simulate the current data, which precludes the climate change studies. Wilks (1998) proposes an interesting multi-site framework using the Richardson model, but it preserves some practical difficulties such as the computational burden and the correlation matrix positive definiteness constraint. Regarding the multi-site generation of temperature and solar radiation data, the main earlier approaches were those of Wilks (1999b) and Buishand and Brandsma (2001). The first one uses an extension of the weakly stationary generating process from 3 dimensions to $3k$ dimensions, where k is the total number of locations. However, this extension yields correlation matrices that are not easy to handle, especially for a large station network and the coefficients of interstation correlations are not necessarily all significant. The second approach uses the historical data, as mentioned above, which cannot deal with climate change studies.

Multi-site generation approach of daily precipitation data has been developed by Khalili *et al.* (2007). This approach adopts the Richardson weather generator and uses the spatial autocorrelation concept to reproduce the spatial dependence exhibited by georeferenced observations. The core idea behind the use of the spatial autocorrelation is its ability to summarize the spatial dependence over a watershed in a single number, which relieves the computation and highlights the straightforwardness of the approach when used for a large data set. Using a proper weight matrix, required for the computation of the spatial

autocorrelation, the multi-site approach can reproduce the observed daily spatial autocorrelations and implies the reproduction of the observed monthly pairwise correlations.

The approach of Khalili *et al.* (2007) allows an accurate simulation for the precipitation processes maintaining the daily spatial autocorrelations and the monthly interstation correlations. However, the simple exponential distribution function used to model the precipitation amounts gives practically good fit to these precipitation amounts, but the spatial intermittence property is not fulfilled. This property must hold because the precipitation amounts at a given station generally depend on whether the surrounding stations are wet or dry. It is thus not appropriate to model the precipitation amounts at a given station without considering the occurrence states in the surrounding stations. Wilks (1998) proposes one attempt to alleviate this issue using a 3-parameter mixed exponential but further improvements are required in this regard. An algorithm simplifying the execution of Wilks' approach has been developed by Brissette *et al.* (2007) with a multi-exponential function to improve the reproduction of seasonal spatial intermittence.

This paper focuses firstly on improving the simulation of the precipitation amounts to obtain an accurate reproduction of daily spatial intermittence using the spatial autocorrelation concept. Secondly, it surveys the use of the spatial autocorrelation for a straightforward multi-site generation of daily temperature and solar radiation data.

Next section presents the general theory of the stochastic multi-site generation approach of daily precipitation data (Khalili *et al.* 2007) and the suggested method to improve the spatial intermittence of the precipitation amounts. Section 4.3 describes the uni-site generation approach of daily temperature and solar radiation data. The proposed methodology for the multi-site generation of these weather data is presented in section 4.4. Section 4.5 presents and discusses the results of the proposed approaches.

4.2 Multi-site generation approach of daily precipitation data

4.2.1 Definition of spatial autocorrelation

Spatial autocorrelation is a correlation between the values of a single variable, taking into account their arrangement in geographic space (Griffith 2003). This statistic complies with the first law of geography mentioned by Tobler (1970) “everything is related to everything else, but near things are more related than distant things”. A more useful spatial autocorrelation statistic is Moran’s I (Moran 1950; Odland 1988; Griffith 2003), presented as:

$$I = \frac{\sum_{i=1}^n (x_i - \bar{x}) \sum_{j=1}^n w_{ij} (x_j - \bar{x}) / \sum_{i=1}^n \sum_{j=1}^n w_{ij}}{\sum_{i=1}^n (x_i - \bar{x})^2 / n} \quad (4.1)$$

where x_i denotes the observed value of a single variable X at location i , \bar{x} is the average of the x_i over n locations and w_{ij} is the spatial weight between two locations i and j . In the matrix form of the right-hand part of Eq. (4.1), all weights are stored in the spatial weight matrix $W(n, n)$. Usually, the weights w_{ij} may be in a row-standardized form, which means that all weights in a row sum up to 1. If the weight matrix is used without row-standardization, this will not change the degree of spatial autocorrelation. In fact, to compensate for the effect of the non row-standardization form, the term $\frac{n}{\sum_{i=1}^n \sum_{j=1}^n w_{ij}}$ in the

Moran’s I expression will not be equal to 1 as it is when each row sums up to 1.

It should be noted that the Moran’s I computes the dependence among values of a single variable, which are weighted according to their localization each one to another. It therefore differs from the Pearson product moment correlation coefficient, which measures the relation between two variables without weight parameters. In precipitation context, the single

variable X used in Moran's I (Eq. (4.1)) represents the values of the precipitation processes observed at the set of stations on a given day such that:

$$X = [x_1 \quad \dots \quad x_n] \quad (4.2)$$

According to the precipitation process to be generated, the x_i may be either the occurrence state at location i or the precipitation amount. For the occurrence process, the value 0 is assigned to the dry state and 1 to the wet state (Khalili *et al.* 2007).

Moran's I has an expected value of $-1/(n-1)$ (Moran 1950; Cliff and Ord 1981) and can vary between -1 and 1 (Griffith 2003). When neighbouring values of a sequenced variable tend to be similar, the resulting Moran's I will be larger than $-1/(n-1)$, and the spatial autocorrelation is said to be positive. When neighbouring values tend to be dissimilar, the resulting Moran's I will be smaller than $-1/(n-1)$, and the spatial autocorrelation is said to be negative. When these values are independent over space, the Moran's I will be equal to $-1/(n-1)$.

The spatial weight matrix is the important concept in the analysis of spatial autocorrelation. This matrix has zero diagonal elements because $w_{ii} = 0$ by convention, and there are no restrictions on the off-diagonal elements. A wide range of suggestions have been proposed in the literature to specify the spatial weights (Odland 1988; Anselin and Bera 1998; Anselin 1980; Murdoch *et al.* 1993). One can select an appropriate spatial weighting function that assigns the weights to every pair of locations. The value of the spatial autocorrelation depends on the selected weight matrix and on the observed data. The weight matrix may not only be derived from the geographical information, but also from the hypotheses about the relation between locations (Odland 1988). Khalili *et al.* (2007), in their multi-site generation approach of daily precipitation data, took advantage from this flexibility in defining the spatial weight between each pair of stations to reproduce not only the observed daily spatial autocorrelations over the watershed, but also the monthly interstation correlations.

4.2.2 Multi-site generation with spatial intermittence approach

The multi-site generation approach of daily precipitation data (Khalili *et al.* 2007) uses a spatial moving average process to generate spatially autocorrelated random numbers whose spatial autocorrelations can reproduce the observed daily spatial autocorrelations in the synthetic time series of precipitation occurrences and amounts. The spatially autocorrelated random numbers are used in the Richardson weather generator to simulate the precipitation occurrences using the first order Markov chain model. The approach can also use the high order of this model.

Another set of spatially autocorrelated random numbers was used to simulate the synthetic precipitation amounts by inverting the simple exponential distribution function. However, to reproduce adequate spatial intermittence, it appears important to think about the relationship between the precipitation amounts at a given station and the occurrence states over the watershed before setting out to model the precipitation amounts. This obviously highlights the potential of employing the spatial autocorrelation to investigate the spatial interaction between the occurrence processes over the watershed. As pointed out, the spatial autocorrelation can be used to summarize the extent of the spatial dependence between the occurrence values at all the stations in a single statistic.

Data from the Peribonca River Basin watershed are used to investigate this purpose. In fact, for a given day, the spatial autocorrelation of precipitation occurrences is computed using equation (4.1). The sequenced variable X (Eq. (4.2)) used in Moran's I represents the values of the precipitation occurrences observed at the set of stations on that day. As anticipated, relationship is found between the Moran values of occurrence processes and the mean precipitation amounts at each station and month. Consequently, the simple exponential distribution will be used, but with the rate parameter, which represents the inverse of the precipitation mean, derived from these relationships. Regression functions can be used to determine the mean precipitation amounts according to the spatial autocorrelations of the occurrence processes computed over the watershed. In doing so, the resulting simple

exponential distribution will be called spatial exponential distribution to signify that the mean of this distribution is defined according to the spatial dependence of the occurrence states at the set of stations.

Furthermore, it may be interesting to suggest another structure of spatial dependence by allowing the Moran's I to be computed without removing the average such that:

$$SDI = \frac{\sum_{i=1}^n x_i \sum_{j=1}^n w_{ij} x_j / \sum_{i=1}^n \sum_{j=1}^n w_{ij}}{\sum_{i=1}^n x_i^2 / n} \quad (4.3)$$

This formula, which provides positive and higher values appears to offer a more significant correlation between the mean precipitation amounts and the spatial dependence of occurrence states. This spatial structure can be called spatial dependence indicator (SDI) to differentiate it from the Moran's I . Of course, this latter is more proper to compute the spatial autocorrelation because it is designed with the measure of covariance. While this condition is logically required for the statistical testing, the multi-site generation approach can be carried out regardless of the data form. One could opt for the formula which deals appropriately with his objective. Note that the Moran's I is not the only indicator that can compute the spatial autocorrelation, but other statistics can be constructed by choosing different measures of covariation (Odland 1988), such as Geary's C (Odland 1988; Griffith 2003).

Thus, to simulate the synthetic precipitation amounts, the multi-site weather generator will use a set of uniform $[0, 1]$ random numbers $v_i(k)$, which are spatially autocorrelated as presented in Khalili *et al.* (2007), but with the inverse of the probability distribution of the spatial exponential function. In fact, the cumulative density function can be defined as:

$$F[r(k)] = 1 - \exp(-\lambda(k)r(k)) \quad (4.4)$$

and

$$r_t(k) = -\ln(1 - v_t(k)) / \lambda_t(k) \quad (4.5)$$

where:

$r_t(k)$ is the synthetic precipitation amount at site k on day t ;

$\lambda_t(k)$ is the inverse of the precipitation mean at site k on day t defined using the regression function established for the given month, and according to the spatial dependence computed for the occurrence values at the set of stations on day t .

4.3 Uni-site generation approach of daily temperature and solar radiation data

To simulate daily maximum temperature, minimum temperature and solar radiation data, Richardson (1981) proposed a weakly stationary generating process (Matalas 1967).

$$\chi_{p,k}(j) = A \chi_{p,k-1}(j) + B \varepsilon_{p,k}(j) \quad (4.6)$$

where:

$\chi_{p,k}(j)$ and $\chi_{p,k-1}(j)$ are (3×1) matrices of maximum temperature ($j=1$), minimum temperature ($j=2$) and solar radiation ($j=3$) residuals for days k and $k-1$ of year p ; Note that these residuals are the deviations of temperatures and solar radiation from the appropriate wet or dry mean $\bar{X}_k(j)$, normalized by the appropriate wet or dry standard deviation $\sigma_k(j)$.

$$\chi_{p,k}(j) = \frac{X_{p,k}(j) - \bar{X}_k(j)}{\sigma_k(j)} \quad (4.7)$$

$\varepsilon_{p,k}(j)$ is (3×1) matrix of independent standard normal random numbers $N[0,1]$ for day k of year p ;

A and B are (3×3) matrices whose elements are defined from lag 0 and lag 1 serial and cross-correlation coefficient matrices of observed residuals.

$$A = M_1 M_0^{-1} \quad (4.8)$$

$$BB^T = M_0 - M_1 M_0^{-1} M_1^T \quad (4.9)$$

where M_0 and M_1 are matrices of lag 0 and lag 1 serial and cross-correlations, respectively, such that:

$$M_0 = \begin{bmatrix} 1 & \rho_{x_{T \max,0} x_{T \min,0}} & \rho_{x_{T \max,0} x_{Sr,0}} \\ \rho_{x_{T \min,0} x_{T \max,0}} & 1 & \rho_{x_{T \min,0} x_{Sr,0}} \\ \rho_{x_{Sr,0} x_{T \max,0}} & \rho_{x_{Sr,0} x_{T \min,0}} & 1 \end{bmatrix} \quad (4.10)$$

$$M_1 = \begin{bmatrix} \rho_{x_{T \max,0} x_{T \max,-1}} & \rho_{x_{T \max,0} x_{T \min,-1}} & \rho_{x_{T \max,0} x_{Sr,-1}} \\ \rho_{x_{T \min,0} x_{T \max,-1}} & \rho_{x_{T \min,0} x_{T \min,-1}} & \rho_{x_{T \min,0} x_{Sr,-1}} \\ \rho_{x_{Sr,0} x_{T \max,-1}} & \rho_{x_{Sr,0} x_{T \min,-1}} & \rho_{x_{Sr,0} x_{Sr,-1}} \end{bmatrix} \quad (4.11)$$

where:

$\rho_{x_{i,0} x_{j,0}}$ is the lag 0 cross-correlation coefficient between the residuals of variable i and the residuals of variable j ;

$\rho_{x_{i,0} x_{j,-1}}$ is the lag 1 cross-correlation coefficient between the current residuals of variable i and the previous residuals of variable j ;

$\rho_{x_{i,0} x_{i,-1}}$ is the lag 1 serial correlation of variable i .

The Cholesky factorisation technique can be used to obtain the B matrix.

Daily maximum temperature, minimum temperature and solar radiation are then found by multiplying the simulated daily residuals by the appropriate wet or dry standard deviation and adding the appropriate wet or dry mean. This of course means that the computation of

daily means and standard deviations is conditioned on the wet or dry state of the day. This state is determined by the simulated occurrence process on that day.

4.4 Multi-site generation approach of daily temperature and solar radiation data

The multi-site generation approach of daily temperature and solar radiation data can be carried out using in the weakly stationary generating process spatially autocorrelated standard normal random numbers. The spatial autocorrelations of these numbers are defined such as to allow the synthetic time series to exhibit the same daily spatial autocorrelations as the observed ones. Note that the observed daily spatial autocorrelations to be reproduced are the averages of the daily spatial autocorrelations observed between the selected weather stations over their shared recording years. As for the precipitation data, this approach focuses on the spatial autocorrelation over the watershed, but the interstation correlations of the weather data should also be automatically reproduced for each pair of stations and month. The Moran's I will serve to compute the spatial autocorrelation. Next section illustrates how to generate spatially autocorrelated random numbers.

4.4.1 Spatially autocorrelated random numbers model

The multi-site generation of daily temperature and solar radiation data deals with the spatial moving average process (Cliff and Ord 1981; Cressie 1993) to generate spatially autocorrelated random numbers such that:

$$V_{T_{\max}} = \gamma_{T_{\max}} \times W \times u_{T_{\max}} + u_{T_{\max}} \quad (4.12)$$

$$V_{T_{\min}} = \gamma_{T_{\min}} \times W \times u_{T_{\min}} + u_{T_{\min}} \quad (4.13)$$

$$V_{Sr} = \gamma_{Sr} \times W \times u_{Sr} + u_{Sr} \quad (4.14)$$

where:

$V_{T_{\max}}(n,1)$, $V_{T_{\min}}(n,1)$ and $V_{Sr}(n,1)$ are vectors of n spatially autocorrelated random numbers to generate the maximum temperature, the minimum temperature and the solar radiation processes respectively at the n locations;

$W(n,n)$ is a weight matrix;

$\gamma_{T_{\max}}$, $\gamma_{T_{\min}}$ and γ_{Sr} are the moving average coefficients to be used for the maximum temperature, the minimum temperature and the solar radiation processes, respectively. The extreme eigenvalues of the weight matrix establish the range of these coefficients, which

is $\left[\frac{-1}{w_{\max}}; \frac{-1}{w_{\min}} \right]$, where w_{\max} is the maximum positive eigenvalue and w_{\min} is the largest

negative eigenvalue in absolute value;

$u_{T_{\max}}(n,1)$, $u_{T_{\min}}(n,1)$ and $u_{Sr}(n,1)$ are vectors of n independent and normally distributed random numbers to be used for the maximum temperature, the minimum temperature and the solar radiation processes, respectively.

The multi-site approach involves the spatial moving average model because it consists of a simple computation that does not require weight matrix operation. Therefore, the multi-site approach remains straightforward for a large data set.

4.4.2 Multi-site temperature and solar radiation model

The multi-site generation approach of daily temperature and solar radiation data aims at using the weakly stationary generating process (Eq. (4.6)) with a vector $\varepsilon_{p,k}(j)$ containing the spatially autocorrelated standard normal random numbers generated by the three earlier spatial moving average processes. For example, to generate the daily temperature and solar radiation data at a 3rd station, the 3rd component from $V_{T_{\max}}$, the 3rd component from $V_{T_{\min}}$ and the 3rd component from V_{Sr} will respectively compose the vector $\varepsilon_{p,k}(j)$ used in the multi-site model such that:

$$\chi_{p,k}(j) = A \chi_{p,k-1}(j) + B \begin{pmatrix} V_{T_{\max}}(3,1) \\ V_{T_{\min}}(3,1) \\ V_{Sr}(3,1) \end{pmatrix} \quad (4.15)$$

Note that the three components $V_{T_{\max}}(3,1)$, $V_{T_{\min}}(3,1)$ and $V_{Sr}(3,1)$ are independent, but each is spatially autocorrelated with the same variables linked to the other stations in the watershed. Also, these components are normally distributed because $u_{T_{\max}}(3,1)$, $u_{T_{\min}}(3,1)$ and $u_{Sr}(3,1)$ are chosen to be normally distributed. However, a mean and variance normalization of $V_{T_{\max}}(3,1)$, $V_{T_{\min}}(3,1)$ and $V_{Sr}(3,1)$ is needed to obtain a mean of zero and a variance of unity required for the weakly stationary generating process. This operation does not affect the spatial autocorrelations of these random numbers. Thus, all the conditions sought by the weakly stationary generating process regarding $\varepsilon_{p,k}(j)$ are satisfied.

As shown by the multi-site model, the weakly stationary generating process is used as in the uni-site model without any further changes. The regionalization of this model is then conducted only by the spatial autocorrelation of the random numbers. The important issue is hence about this spatial autocorrelation. In fact, for a given weight matrix and for different values of the moving average coefficients $\gamma_{T_{\max}}$, $\gamma_{T_{\min}}$ and γ_{Sr} , the spatial moving average processes can provide spatially autocorrelated random numbers with different degrees of spatial autocorrelation. It is then worth asking which values of $\gamma_{T_{\max}}$, $\gamma_{T_{\min}}$ and γ_{Sr} should be used in the spatial moving average processes to reproduce the observed spatial autocorrelations in the synthetic daily temperature and solar radiation data.

4.4.3 Moving average coefficients estimation

Obtaining suitable values of $\gamma_{T_{\max}}$, $\gamma_{T_{\min}}$ and γ_{Sr} is realized in the same manner as with the precipitation data (Khalili *et al.* 2007). In fact, using different values of these coefficients from their range defined above, one can obtain via the three spatial moving average processes (Eqs. (4.12), (4.13) and (4.14)) spatially autocorrelated random numbers, which

exhibit different spatial autocorrelation values noted $I_{V,T_{\max}}$, $I_{V,T_{\min}}$ and I_{V,S_r} . These notations mean the spatial autocorrelations computed by the Moran's I (Eq. (4.1)) for $V_{T_{\max}}$, $V_{T_{\min}}$ and V_{S_r} intended for the generation of maximum temperature, minimum temperature and solar radiation, respectively. When these random numbers are integrated in the weakly stationary generating process, it generates synthetic time series that also have different spatial autocorrelations noted $I_{T_{\max}}$, $I_{T_{\min}}$ and I_{S_r} .

Therefore, this operation can give way to three relationships. The first one is between the coefficients $\gamma_{T_{\max}}$ and the resulting spatial autocorrelations $I_{T_{\max}}$ for the synthetic maximum temperature data. The second one is between the coefficients $\gamma_{T_{\min}}$ and the resulting spatial autocorrelations $I_{T_{\min}}$ for the synthetic minimum temperature data, and the third relationship is assumed to be established between the coefficients γ_{S_r} and the resulting spatial autocorrelations I_{S_r} for the synthetic solar radiation data. These relationships $(I_{T_{\max}}, \gamma_{T_{\max}})$, $(I_{T_{\min}}, \gamma_{T_{\min}})$ and (I_{S_r}, γ_{S_r}) can be established monthly and used to extract the $\gamma_{T_{\max}}$, $\gamma_{T_{\min}}$ and γ_{S_r} values linked to the observed $I_{T_{\max}}$, $I_{T_{\min}}$ and I_{S_r} , respectively.

The spatial moving average processes will then be reused with the appropriate $\gamma_{T_{\max}}$, $\gamma_{T_{\min}}$ and γ_{S_r} values to provide the spatially autocorrelated random numbers that allow the reproduction of the observed $I_{T_{\max}}$, $I_{T_{\min}}$ and I_{S_r} in the synthetic temperature and solar radiation time series.

Figure 4.1 recapitulates the proposed multi-site generation approach of daily temperature and solar radiation data. The parameter estimation and generation phases are presented for a given month. m is the total number of $\gamma_{T_{\max}}$, $\gamma_{T_{\min}}$ and γ_{S_r} values taken from their range and l is the total number of days in a given month.

4.5 Results and discussion

4.5.1 Results of spatial intermittence approach

The multi-site generation approach of daily precipitation data (Khalili *et al.* 2007) was tested with data from the Peribonca River Basin (and surrounding area) (Figure 4.2). This region, which is characterized by a wet climate, relatively cool summers and snow precipitations from November to April, is also used to test the performance of the approaches proposed in this paper. Therefore, the seven stations in this watershed are selected: Peribonca (P), Hemon (H), Bonnard (B), Chute-du-Diable (Cd), Chute-des-Passes (Cp), St-Léon-de-Labrecque (Sl) and Normandin CDA (N). The location of these stations is provided in Figure 4.2 and in Table 4.1.

As was argued before, the precipitation amounts at each station should be simulated taking into account the occurrence spatial dependence over the watershed. To carry out this method, one first needs to select a spatial weight matrix, which represents the relation between each pair of stations. For this issue, the inverse square distance method is used to represent the distance relations among the weather stations but other spatial weighting functions not limited to the geographical information may be used. The resulting weight matrix is:

$$W = \begin{bmatrix} 0 & 0.0980 & 0.0053 & 0.4972 & 0.0154 & 0.1963 & 0.1879 \\ 0.1426 & 0 & 0.0087 & 0.0745 & 0.0226 & 0.0511 & 0.7006 \\ 0.0956 & 0.1082 & 0 & 0.0997 & 0.5103 & 0.0946 & 0.0916 \\ 0.2707 & 0.0279 & 0.0030 & 0 & 0.0094 & 0.6494 & 0.0396 \\ 0.1462 & 0.1475 & 0.2677 & 0.1646 & 0 & 0.1523 & 0.1217 \\ 0.1314 & 0.0235 & 0.0035 & 0.7986 & 0.0107 & 0 & 0.0321 \\ 0.2325 & 0.5960 & 0.0063 & 0.0901 & 0.0158 & 0.0594 & 0 \end{bmatrix} \quad (4.16)$$

The first row contains the spatial weights between the Peribonca station and the other stations in the order given above. The second row presents the spatial weights between the Hemon station and the remaining stations, and so on. Note that the spatial weight between each station and itself is 0 by convention and the matrix is row-standardized.

Using the spatial dependence indicator (*SDI*) discussed before, the spatial dependence of precipitation occurrences were computed using the shared period for all stations, which is 14 years starting from 1963 to 1976. Relationships are then obtained between the mean precipitation amounts and the *SDI* values for each station and month. Figure (4.3) shows an example of such relationship obtained for Peribonca station in September. A regression fit is used to evaluate the mean precipitation amounts according to the spatial dependence of occurrence values.

As was introduced by Wilks (1998), the statistic “continuity ratio” is used to test the accuracy of the spatial intermittence in the synthetic precipitation time series. This statistic is computed for each pair of station (i, j) and is the ratio of the mean of the nonzero precipitation amounts at the station i when the station j is dry, to the mean of the nonzero precipitation amounts at the station i when the station j is wet, such that:

$$continuity\ ratio = \frac{E[r_t(i) | r_t(i) > 0, r_t(j) = 0]}{E[r_t(i) | r_t(i) > 0, r_t(j) > 0]} \quad (4.17)$$

Figures 4.4a and 4.4b illustrate the spatial intermittence results using the simple exponential function (a) and the spatial exponential functions (b) to fit the precipitation amounts. The use of the spatial exponential functions greatly improves the results and confirms the dependence of the precipitation means at a given station on the occurrence states at the set of stations. Each point in this graph represents the continuity ratio for each pair of stations and month. 504 points are therefore plotted in this graph. The departure from the 45-degree reference line can be attributed to the scatter plot obtained in some relationships between the mean precipitation amounts and the *SDI* values of the occurrences due to sampling variations related to the small sample size (14 years).

4.5.2 Results of multi-site generation approach of daily temperature data

Multi-site generation approach of daily temperature data is also tested using the Peribonca River Basin watershed. The multi-site model is used only for these weather data because solar radiation data are not available for these stations. The results presented are for a 50-years simulation period.

As was the case for the precipitation processes (Khalili *et al.* 2007), the multi-site generation approach of daily temperature data was achieved using different weight matrices. In a first step, the results for the row-standardized weight matrix using the inverse square distance method, presented above, are reported.

Using this weight matrix, the variation range of the moving average coefficients $\gamma_{T_{\max}}$ and $\gamma_{T_{\min}}$ is $]-1; 1.3254[$, according to the extreme eigenvalues of this matrix. Therefore, taking values from this range, the synthetic temperature data exhibit spatial autocorrelations $I_{T_{\max}}$ and $I_{T_{\min}}$ which are linked to these moving average coefficient values. Figure 4.5 shows such relationship for maximum temperature in July. Similar curves are obtained for the other months and also for minimum temperature. Because the third-order polynomial functions produce the best fit to a data set, they may be chosen to describe these relationships.

Once the $\gamma_{T_{\max}}$ and $\gamma_{T_{\min}}$ values linked to all the observed $I_{T_{\max}}$ and $I_{T_{\min}}$ are found for each month, the spatial moving average processes are reused as well as the multi-site temperature model to generate synthetic time series with the required values of the daily spatial autocorrelations.

Figures 4.6a and 4.6b illustrate the observed and simulated daily spatial autocorrelations of maximum and minimum temperature, respectively. The observed ones are averaged over the shared recorded period between the seven stations and the simulated ones are averaged over the 50-years simulation period. While the daily spatial autocorrelations are adequately

reproduced, the monthly interstation correlations are underestimated as shown by figures 4.7a and 4.7b. The observed ones are calculated over the shared recorded period between each pair of stations and the simulated ones are calculated over the simulation period. Note that the observed monthly pairwise correlations are assumed to be reproduced in the synthetic daily temperature data automatically via the reproduction of the observed daily spatial autocorrelations.

This problem occurred because the used weight matrix is based on the inverse square distance method. Such a distance criterion can result in spatial weights, which are practically null. Therefore, a loss of correlation will be observed between the corresponding pair of stations, which are actually correlated. It is thus much important to not mechanically use the distance criterion methods to model the weight matrix, but rather to select a proper function that adequately describes the dependence between each pair of stations, especially for this multi-site approach, which has to satisfy the interaction over the watershed.

Khalili *et al.* (2007) returned to this issue in the multi-site generation approach of daily precipitation data and they mitigated this problem using monthly weight matrices modeled with monthly interstation correlations of daily precipitation data. This paper pursues the same idea for the temperature data but nothing limits the spatial weights to this meaning. Therefore, another analysis is realized using monthly weight matrices expressed by the monthly interstation correlations of daily temperature data.

However, the analyses conducted for this study dealing with the Moran's I reveal a particular behaviour of this statistic when the georeferenced data are practically similar or near to their average. In fact, the Moran's I measures the spatial autocovariance of neighbouring values composing a single sequenced variable. Thus, if all values are close to their average, the resulting Moran's I will be equal to its expected value $-1/(n-1)$. Of course, this does not mean that the values are independent over space, but rather, it is a consequence of the similarity of the values with their average.

This behaviour can occur here when the n values t of daily maximum or minimum temperature observed at the n stations are practically close to their average. Therefore, the Moran's I of a variable T containing the n daily temperature values such that:

$$T = [t(1) \quad \dots \quad t(n)] \quad (4.18)$$

will be around its expected value.

However, this issue should not be regarded as a problem. On the contrary, because in this case the values of spatial autocorrelation are approximately equal to the Moran's I expectation, one does not need to establish the monthly relationships between the moving average coefficients and the spatial autocorrelations of both maximum and minimum temperatures to reproduce the observed daily spatial autocorrelations. One can simply use the spatial moving average processes with $\gamma_{T_{\max}}$ and $\gamma_{T_{\min}}$ of unity or equal to the mean over their range.

Using the monthly weight matrices, adequate interstation correlations were computed for the synthetic daily temperature data for each pair of stations and month as shown by Figures 4.8a and 4.8b. Figures 4.9a and 4.9b illustrate the observed versus simulated daily spatial autocorrelations of maximum and minimum temperature data, respectively. Some values of spatial autocorrelations are underestimated, but generally, this figure shows a good agreement between the observed and simulated daily spatial autocorrelations. Figures 4.10a and 4.10b illustrate the observed versus simulated monthly means of daily maximum and minimum temperature, respectively, for each station. The observed data are averaged over the recorded period for each station and the simulated ones are averaged over the simulated period. It seems that the multi-site generation approach adequately reproduces these statistics.

Because the seven stations are practically all close to each other, all the interstation correlations of daily temperature data are high. It is thus interesting to check whether the

multi-site generation approach can reproduce different magnitudes of monthly pairwise correlations. Two other stations, Nitchequon and Shefferville, located north of the Peribonca watershed are thus added to this analysis. Because these two stations are located far from the others, they are unlikely to be influenced by the same weather phenomena and the monthly correlations between them and the remaining stations are not high.

Figures 4.11a and 4.11b illustrate the observed and simulated interstation correlations of daily maximum and minimum temperature, respectively, between each pair of the nine stations and month. It appears that the multi-site approach adequately reproduces different values of correlations. The daily spatial autocorrelations are also reproduced as shown by Figures 4.12a and 4.12b. The simulated values are in accordance with the observed ones.

The same multi-site generation approach is also applied using the *SDI* instead of Moran's *I*. Figures 4.13a and 4.13b show the observed and simulated interstation correlations of daily maximum and minimum temperature, respectively, between each pair of the nine stations and month. The results are generally satisfactory despite the overestimation of some correlations in the minimum temperature case. The daily *SDI* values are also efficiently reproduced as shown by figures 4.14a and 4.14b.

4.6 Conclusion

Multi-site generation approach using spatial autocorrelation has been proposed. This concept summarizes the spatial dependence over the entire watershed in a single statistic. The essence of this approach is to reproduce at once the observed daily spatial autocorrelations and the observed monthly interstation correlations using proper spatial weight matrices. The basic idea is to use in the weather generator suitable spatially autocorrelated random numbers, which can reproduce these observed statistics.

This paper is an extension of the multi-site generation approach proposed by Khalili *et al.* (2007). A first issue was the improvement of the synthetic spatial intermittence. The feasible

solution presented here is to approximate the mean precipitation amounts used in the simple exponential function by a regression function of the occurrence spatial dependence at the set of stations. Such a method was proposed because the mean precipitation amounts at a given weather station is not a free parameter, but depend on the occurrence states over the watershed. Therefore, monthly spatial exponential functions were established for each station. This approach greatly improves the reproduction of the synthetic spatial intermittence.

A multi-site generation approach of daily maximum temperature, minimum temperature and solar radiation data was also presented in this paper. This approach adopted the weakly stationary generating process used in the Richardson weather generator, with no modifications made, but only the random numbers have to be spatially autocorrelated. Spatial moving average processes were used to generate spatially autocorrelated random numbers that can reproduce the observed daily spatial autocorrelations in the synthetic time series. Because of lack of solar radiation data, the multi-site weakly stationary generating process was only established for the temperature data. This approach has been shown to perform well in the generation of these data. Sufficiently accurate daily spatial autocorrelations and monthly interstation correlations were obtained in the synthetic time series. Some aberrations were produced generally due to the small sample size.

The suggested spatial dependence indicator using raw data gives also adequate results. One can use this indicator, especially when the data of interest are practically similar and near to their average.

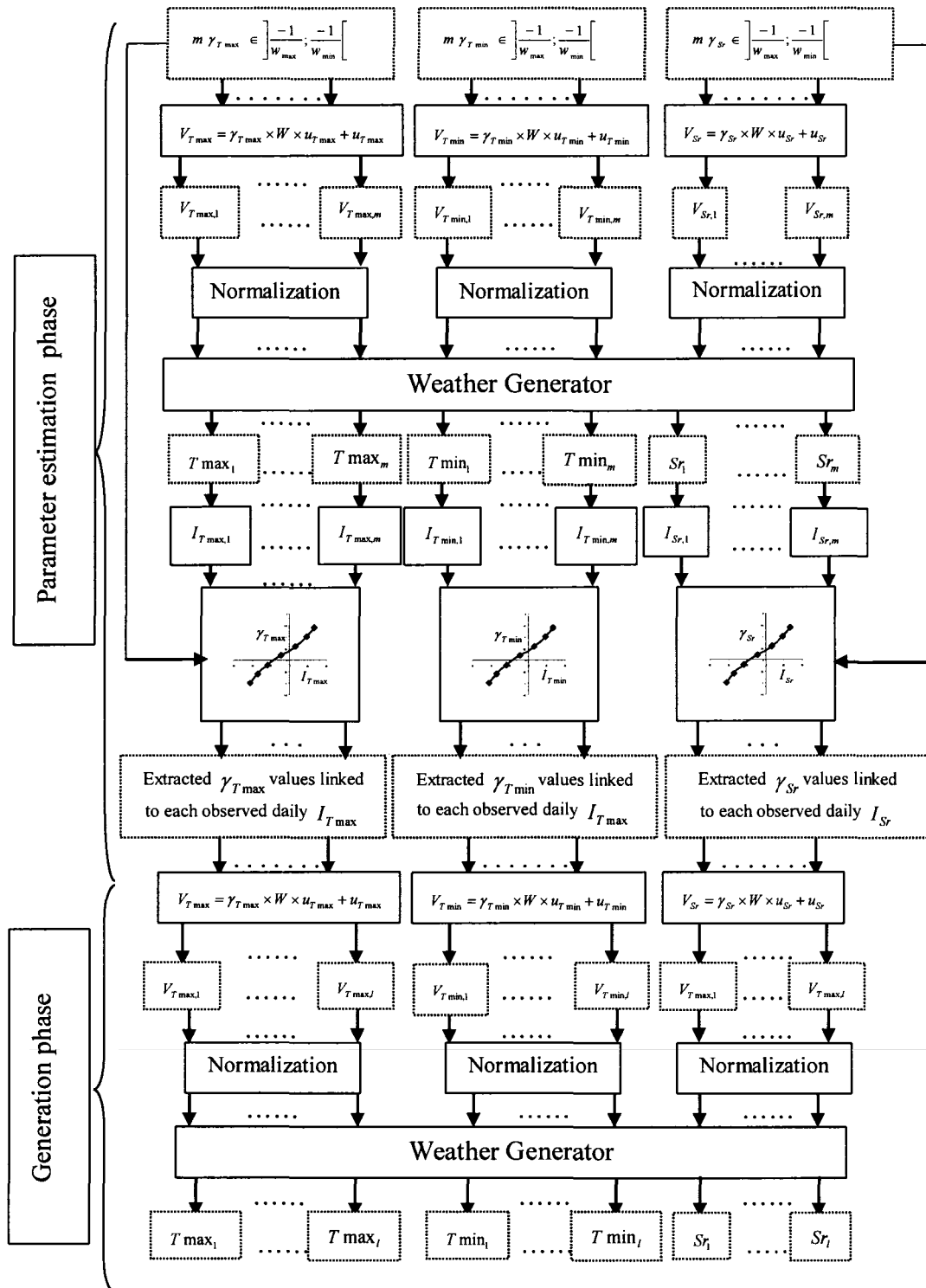


Figure 4.1 Flow chart for the proposed multi-site generation approach of daily temperature and solar radiation data.

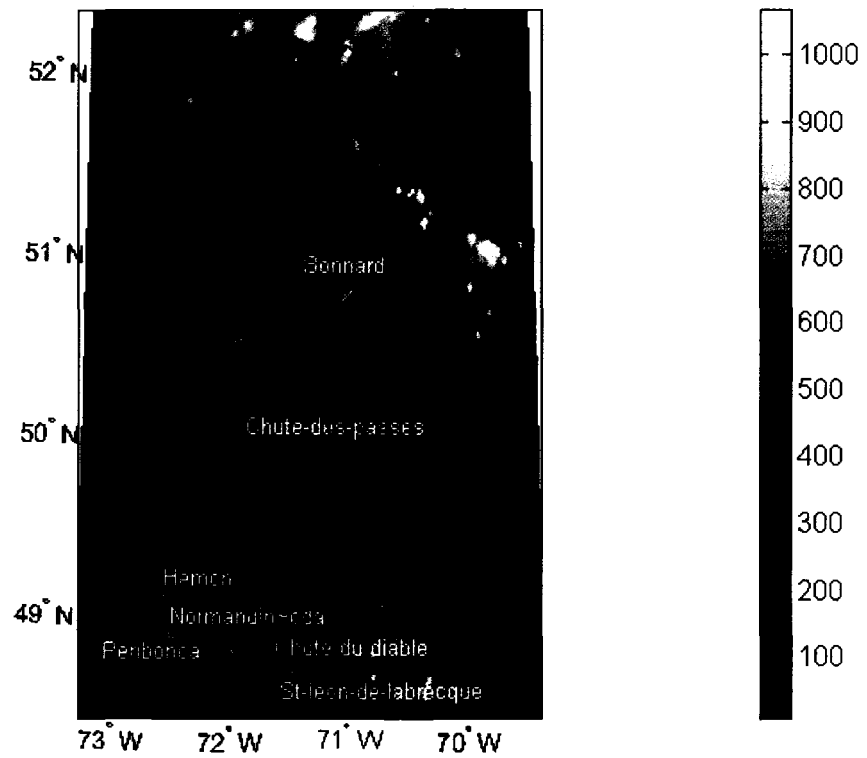


Figure 4.2 *The Peribonca watershed, showing the location of meteorological stations. Elevations in meters.*

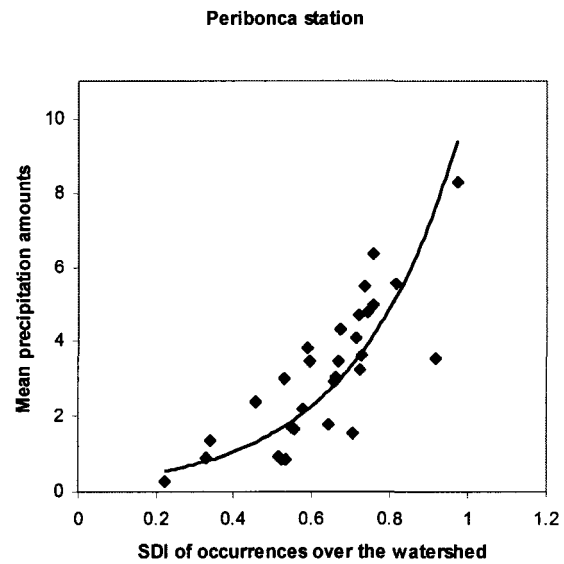


Figure 4.3 *Relationship between the mean precipitation amounts (mm) of Peribonca station and the spatial dependence values of occurrences over the watershed in September.*

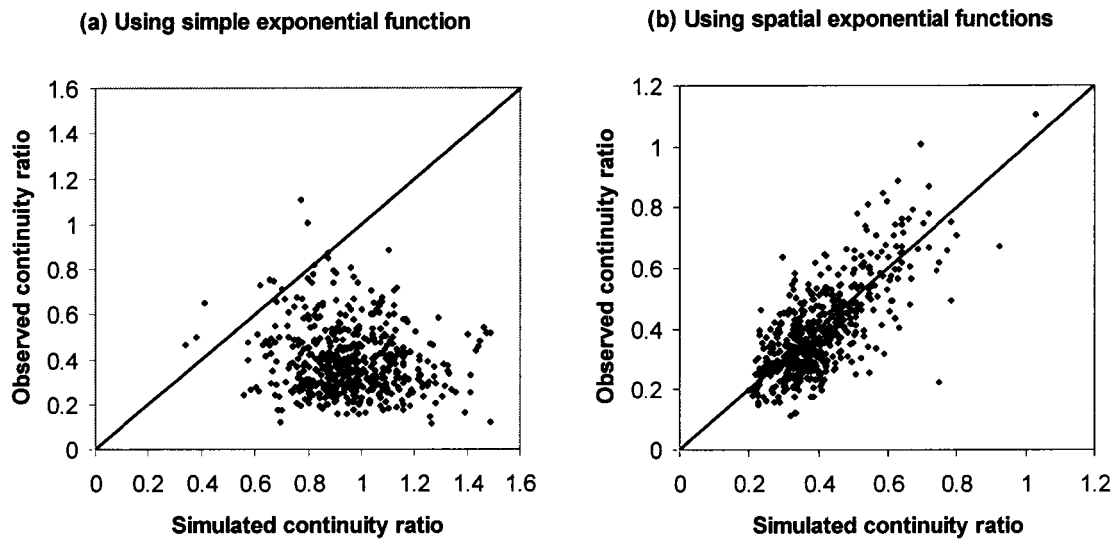


Figure 4.4 *Spatial intermittence using the simple exponential function (a) and the spatial exponential functions (b).*

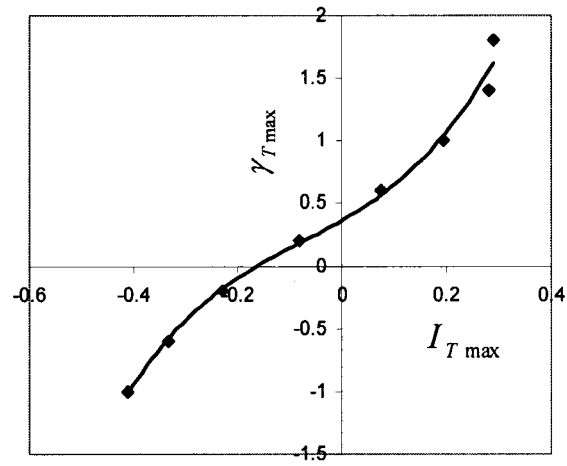


Figure 4.5 *Relationship between the moving average coefficients $\gamma_{T \max}$ and the spatial autocorrelations $I_{T \max}$ for maximum temperature data in July.*

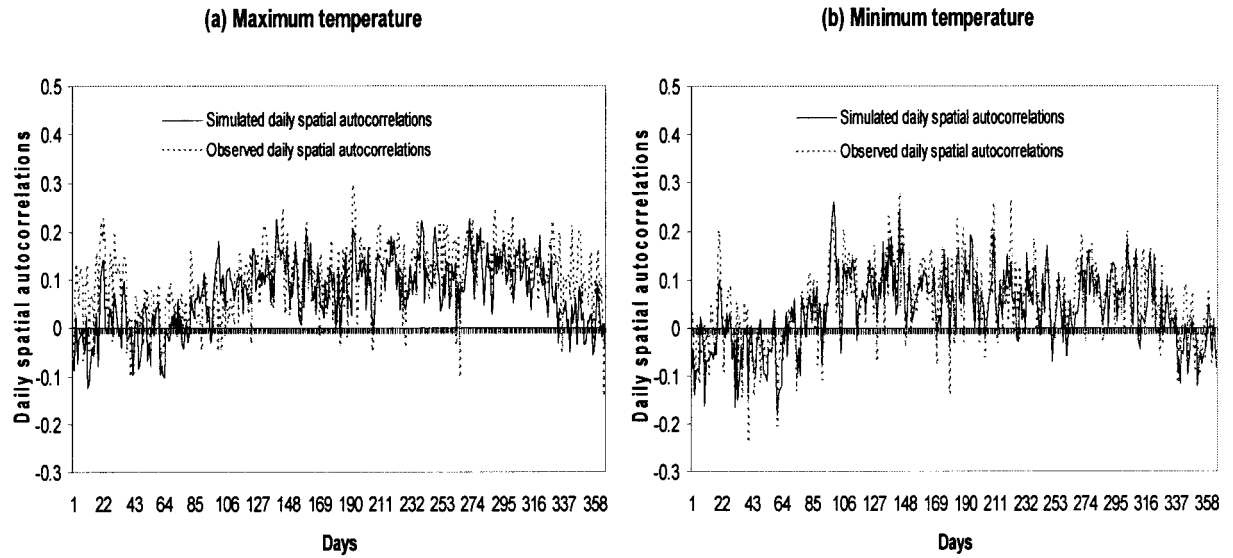


Figure 4.6 *Observed and simulated daily spatial autocorrelations of maximum temperature (a) and minimum temperature (b).*

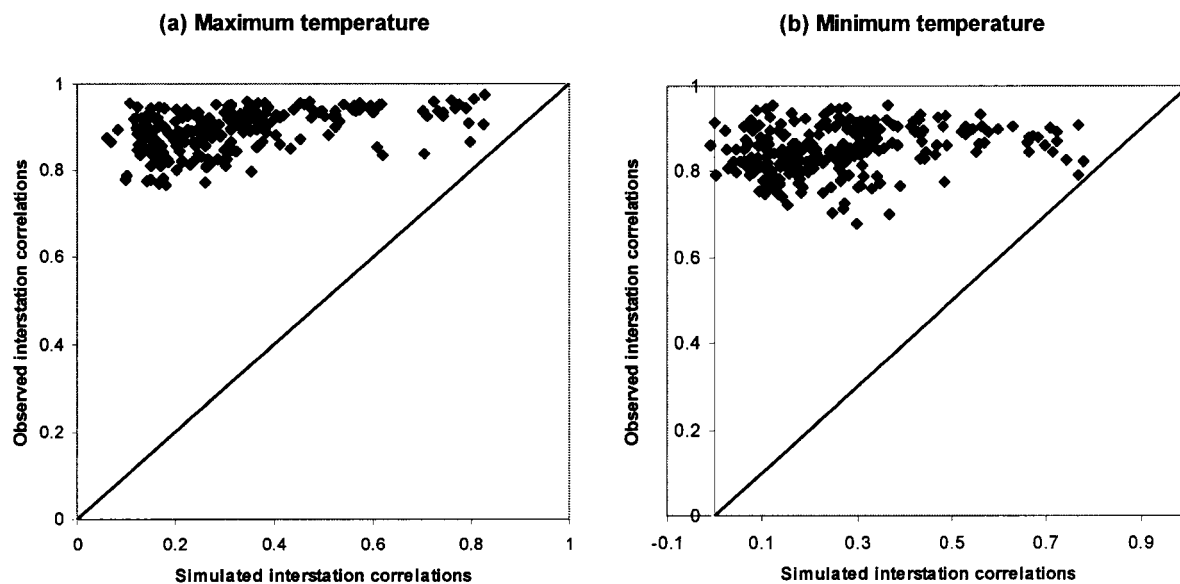


Figure 4.7 *Observed versus simulated interstation correlations of daily maximum temperature (a) and minimum temperature (b) between each pair of the seven stations and month.*

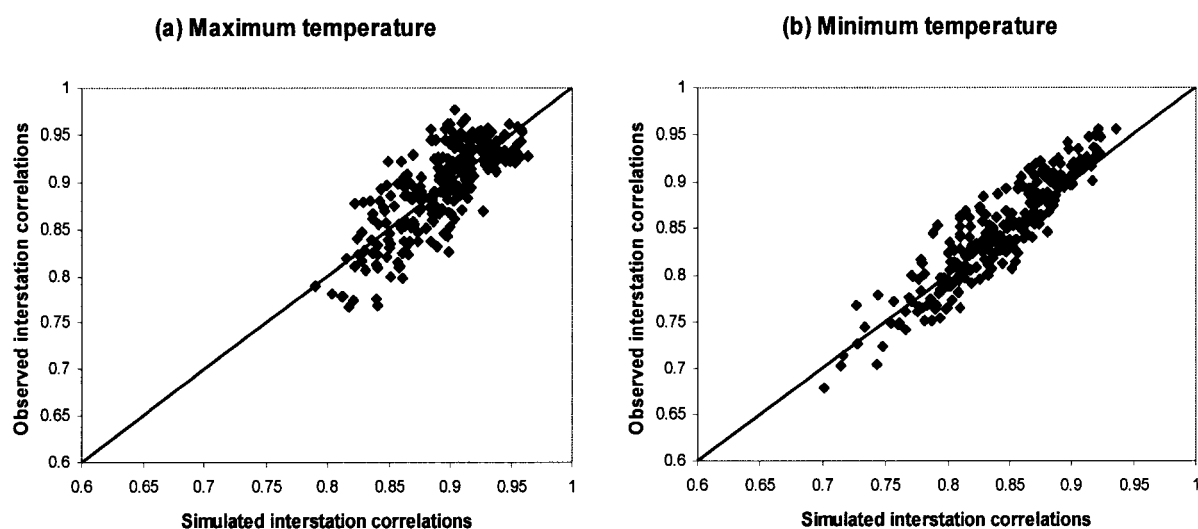


Figure 4.8 *Observed versus simulated interstation correlations of daily maximum temperature (a) and minimum temperature (b) between each pair of the seven stations and month using the monthly weight matrices.*

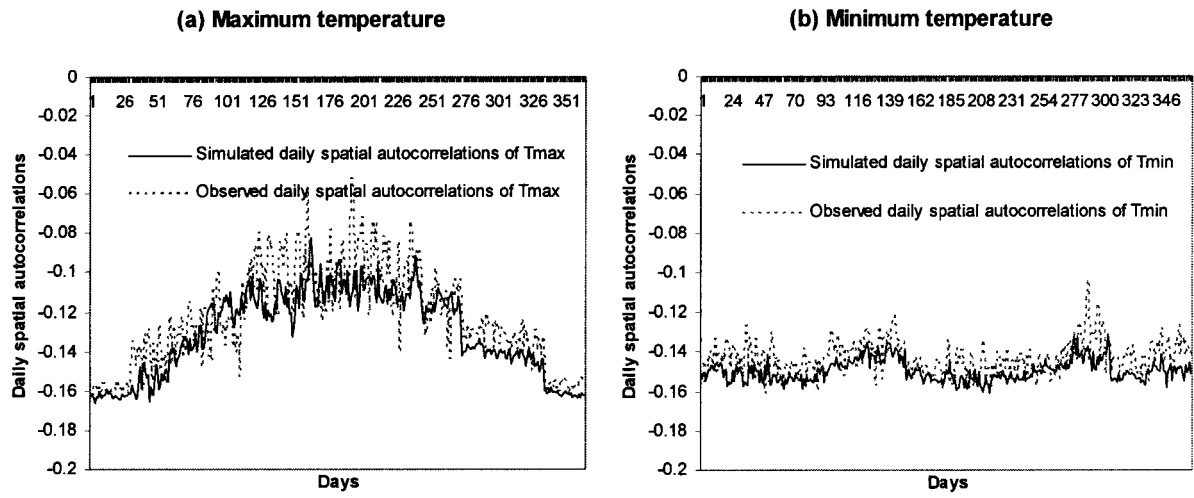


Figure 4.9 *Observed and simulated daily spatial autocorrelations of maximum temperature (a) and minimum temperature (b) using the monthly weight matrices.*

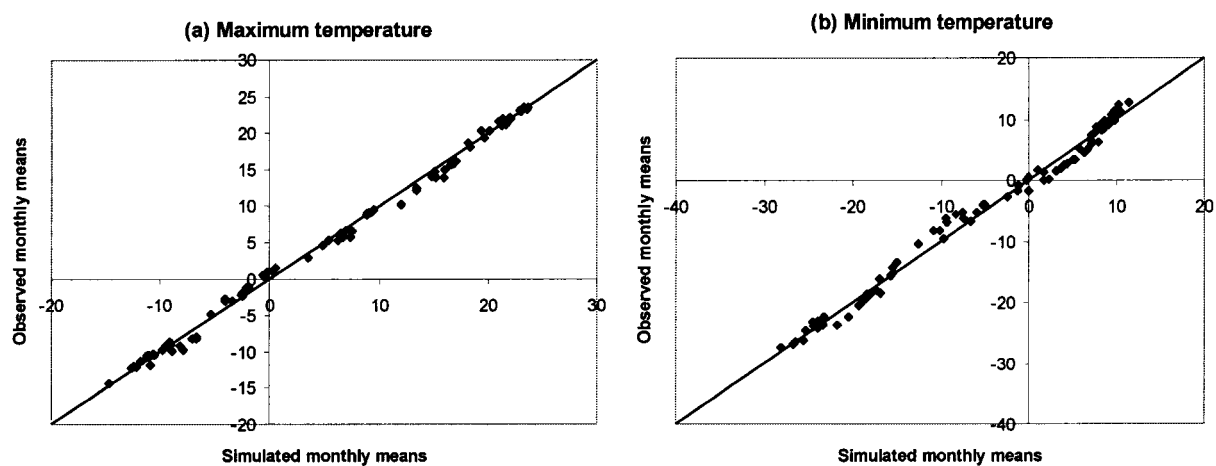


Figure 4.10 *Observed versus simulated monthly means of daily maximum temperature (a) and minimum temperature (b).*

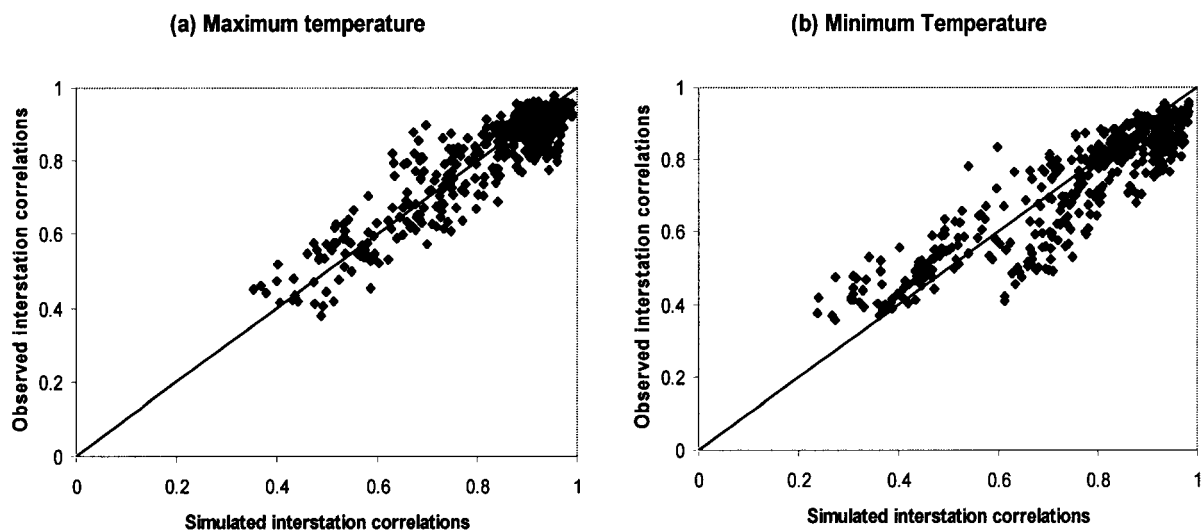


Figure 4.11 *Observed versus simulated interstation correlations of daily maximum temperature (a) and minimum temperature (b) between each pair of the nine stations and month using the monthly weight matrices.*

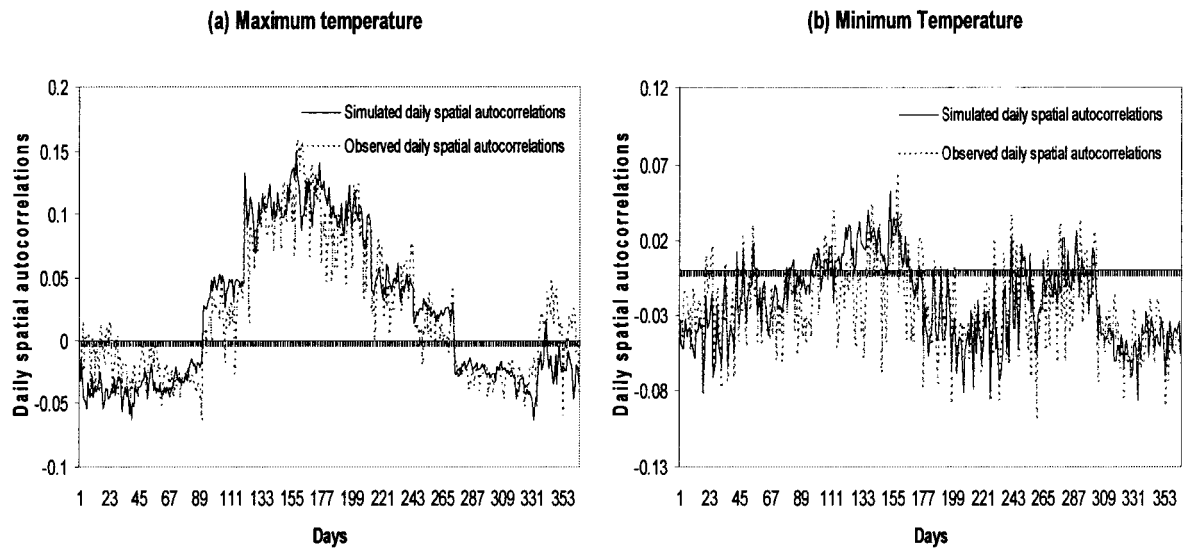


Figure 4.12 *Observed and simulated daily spatial autocorrelations of maximum temperature (a) and minimum temperature (b) using the nine stations and the monthly weight matrices.*

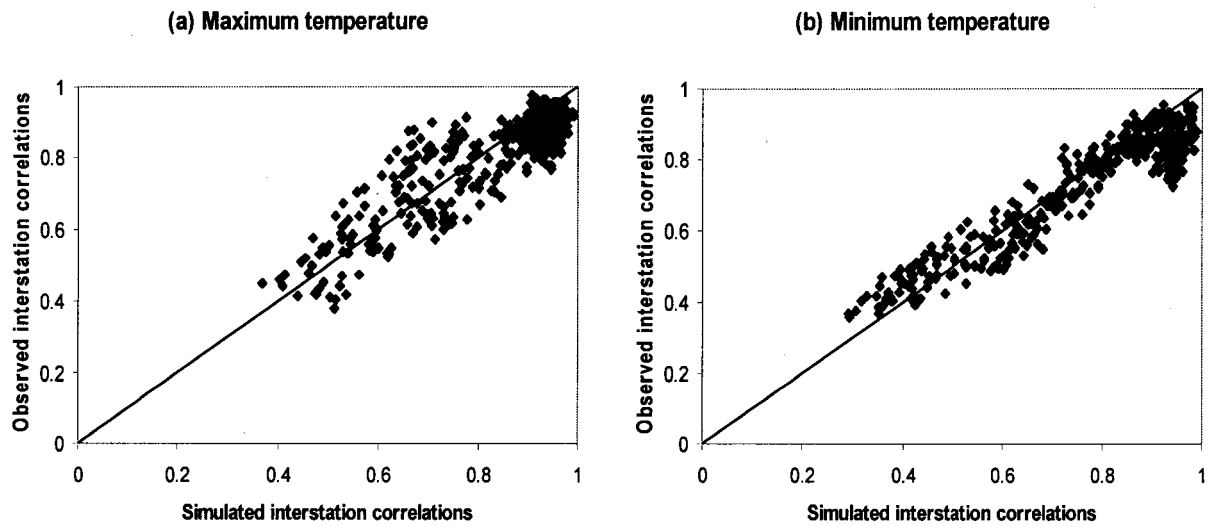


Figure 4.13 *Observed versus simulated interstation correlations of daily maximum temperature (a) and minimum temperature (b) between each pair of the nine stations and month using the monthly weight matrices and SDI.*

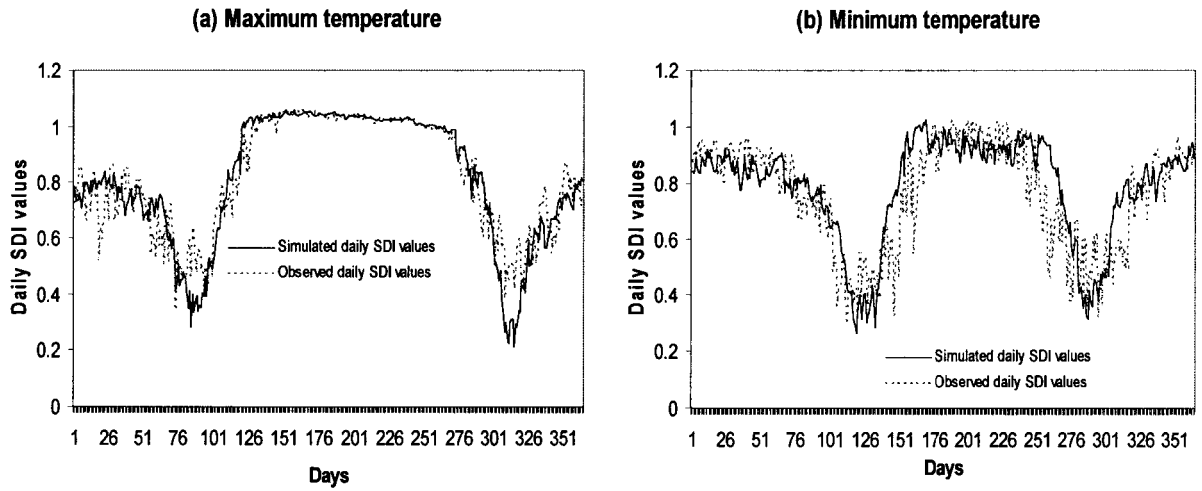


Figure 4.14 *Observed and simulated daily spatial dependence of maximum temperature (a) and minimum temperature (b) using the monthly weight matrices and SDI.*

Tableau 4.1

Location and recorded years of the used nine stations

Stations	Lat	Long	Elevation (m)	Recorded years
Péribonca	48°45'42"N	72°01'24"W	107	1951-2002
Hémon	49°03'48"N	72°35'35"W	206	1963-2002
Chute-du-Diable	48°45'00"N	71°42'00"W	174	1951-1976
St-Léon de Labrecque	48°40'26"N	71°31'33"W	131	1963-1997
Normandin CDA	48°51'00"N	72°32'00"W	137	1936-1992
Chute-des-Passes	49°50'25"N	71°10'06"W	398	1960-1976
Bonnard	50°44'00"N	71°02'00"W	506	1961-2000
Nitchequon	53°12'N	70°54'W	-----	1960-1979
Shefferville	54°48'N	66°49'W	-----	1960-1979

CHAPITRE 5

ARTICLE 3: EFFECTIVENESS OF MULTI-SITE WEATHER GENERATOR FOR HYDROLOGICAL MODELLING: COMPARATIVE STUDY WITH UNI-SITE WEATHER GENERATOR USING DISTRIBUTED HYDROLOGICAL MODEL

Abstract

A multi-site weather generator has been developed using the concept of spatial autocorrelation. This concept describes the spatial dependence over the entire watershed with a single number. The multi-site generation approach reproduces the spatial autocorrelations observed between a set of weather stations as well as the correlations between each pair of stations, using a proper weight matrix. Its performance has been assessed in two previous studies using both precipitation and temperature data. The main objective of this paper is to assess the efficiency of this multi-site weather generator compared to a uni-site generator with respect to hydrological modelling.

A hydrological model, known as Hydrotel, was applied over the Chute du Diable watershed, located in the Canadian province of Quebec. The distributed nature of Hydrotel accounts for the spatial variations throughout the watershed, and thus allows a more in-depth assessment of the effect of spatially dependent meteorological input on runoff generation. After the model was calibrated and validated, simulated streamflows using both multi-site and uni-site generated weather data were statistically compared to flows modelled using observed data.

Overall, the multi-site weather generator significantly outperformed the uni-site generator. A frequency analysis revealed a significant underestimation of extreme streamflows for all seasons, when the uni-site weather generator was used. More consistent results were obtained using the multi-site generated weather data. Monthly streamflow averages and standard deviations were also satisfactorily reproduced using multi-site weather generated data. Discrepancies were observed with the uni-site weather generator due to a low variability of

uni-site simulated air temperature data. This variability was better reproduced using the multi-site weather generator, but more improvement is required for the winter season.

Keywords: Weather generator; precipitation; temperature; multi-site; hydrological modelling; Hydrotel.

5.1 Introduction

The evaluation of the effects of climate change on river hydrological regimes using hydrological models coupled to climate projections is one of the issues covered by climate change studies. Engineering design, water resource management, water supply and water quality studies also require hydrological watershed modelling. Meteorological data is used as the primary input for hydrological models, and various applications, including extreme event analyses, require long time series of meteorological data, which unfortunately, are not always readily available. Long-term meteorological data are also required in conducting climate change impact studies.

Stochastic weather generators were used to produce meteorological data time series for practically any time period. The statistical properties of the simulated time series are similar to those of the observed ones. Weather generators are particularly useful for climate change studies, and using climate change projections derived from Global Circulation Models (GCMs), the weather generator parameters can be “perturbed” to simulate future time series.

The WGEN weather generator (Richardson, 1981; Richardson and Wright, 1984) and the Long-Ashton Research Station Weather Generator (LARS-WG) (Semenov and Barrow, 1997) are well known uni-site parametric weather generators. WGEN uses the first order two-state Markov chain model to generate precipitation occurrences and a selected distribution function to model precipitation amounts on rainy days, while the LARS-WG uses semi-empirical distributions to model precipitation occurrences and amounts. The weather generator developed by Brandsma and Buishand (1997) is a non-parametric uni-site generator, which uses the resampling of a weather variables vector on a day of interest from the historical data by conditioning on the simulated values of previous days.

Multi-site weather generators have been developed to take into account the correlations existing between weather stations, a feature which cannot be addressed by uni-site generators. The literature contains models, which use the atmospheric circulation patterns,

such as the space-time models (Bardossy and Plate, 1992; Bogardi *et al.* 1993), and the non-homogeneous hidden Markov models (Bellone *et al.*, 2000; Hughes and Guttorp, 1994a, 1994b; Hughes *et al.*, 1999). However, these models are complex, and do not reproduce an adequate spatial dependence between the weather stations. Buishand and Brandsma (2001) regionalised the uni-site weather generator cited above (Brandsma and Buishand, 1997), but because of the nearest-neighbour resampling from historical data it requires, this weather generator is not appropriate for climate change studies. Wilks (1998) regionalised WGEN using serially independent but spatially correlated random numbers. This multi-site weather generator produces adequate results, but it also has a cumbersome structure involving a lot of parameters, such as the collection of $k(k-1)/2$ empirical relationships for a network of k stations, relating the correlations of the random numbers and the correlations of occurrence and amount precipitations. These curves must be developed for all possible pairs of stations and for each month.

The multi-site weather generators presented above focus mainly on precipitation processes, which are precipitation occurrences and amounts. Very few multi-site models include other meteorological variables. Wilks (1999b) presented an extension of the weakly stationary generating process used in WGEN for the multi-site generation of maximum temperature, minimum temperature and solar radiation data. However, this method, which extends the model dimension from 3 to $3k$, significantly increases the size of correlation matrices, and thus precludes the model solution. The Buishand and Brandsma (2001) multi-site weather generator uses the resampling from the historic record for both precipitation and temperature data, but the method does not permit the generation of time series with climate change.

A multi-site generation approach for daily precipitation and temperature data has been developed by Khalili *et al.* (2007, 2008a) using the spatial autocorrelation concept. The interest for this concept lies in its ability to describe the spatial dependence between neighbouring sites over the entire watershed with a single number. The multi-site generation approach proved successful when applied for simulating daily weather data. Using a proper weight matrix to compute the spatial autocorrelation allowed the reproduction of the daily

spatial autocorrelations between the set of weather stations as well as the monthly interstation correlations between each pair of stations.

According to Wilks and Wilby (1999), Mehrotra *et al.* (2005), Srikanthan and McMahon (2001) and Harmel *et al.* (2002), ignoring the spatial correlation between weather stations using uni-site weather generators in hydrological modelling studies can introduce significant errors to model results. However, this assertion has not been established in practice as major studies usually tend to evaluate hydrological modelling using uni-site weather generators (Dubrovsky *et al.*, 2004; Harmel *et al.*, 2000; Siriwardena *et al.*, 2002).

Very few researches have actually investigated the performance and usefulness of multi-site weather generators for hydrological modelling. Waston *et al.* (2005) evaluated the response of the semi-distributed hydrological model Soil and Water Assessment Tool (SWAT) to uni-site and multi-site weather generators (Wilks, 1998) over the Woody Yaloak River catchment (306 km²) in Australia. The authors report little distinction between the two weather generators based on the means and standard deviations of annual, monthly and daily runoff, and recommend more studies to further assess the usefulness of multi-site weather generators in hydrological modelling. Because Woody Yaloak is a small watershed (306 km²), it was somewhat expected that a uni-site weather generator would give results comparable to those of a multi-site generator.

Khalili *et al.* (2006) recently carried out a comparison between WGEN and multi-site weather generator (Khalili *et al.*, 2007) using a lumped conceptual hydrological model (Bisson and Roberge, 1983). The multi-site and uni-site simulated precipitation time series were spatially averaged over the watershed and used as input to the lumped model. One interesting observation was that although the precipitation data was averaged over the basin, some spatial dependence was retained in the averaging process, which resulted in an improved summer-autumn flow modelling for the multi-site approach.

This article thus aims to compare the performance of a multi-site weather generator (Khalili *et al.*, 2007, 2008a) against the uni-site WGEN weather generator in modelling the hydrological regime of the Chute du Diable watershed in the Canadian province of Quebec. This study uses a distributed hydrological model called Hydrotel (Fortin *et al.*, 2001a, 2001b). Using a distributed modelling approach provides greater insights into the performance of each multi-site and uni-site weather generator. The uni-site weather generator is described in the next section, while section 5.3 presents the multi-site weather generator (Khalili *et al.*, 2007, 2008a) for daily precipitation and other climate data. Section 5.4 presents the hydrological modelling, and section 5.5 shows and discusses the results of the comparison.

5.2 Uni-site weather generator

5.2.1 Uni-site generation of daily precipitation data

The uni-site weather generator used for this study is WGEN (Richardson, 1981; Richardson and Wright, 1984). It uses a first-order two-state Markov chain to simulate the daily precipitation occurrence $x_t(i)$ at site i on day t . The operation consists in computing two transitional probabilities and selecting one according to the occurrence state of the previous day. The selected transitional probability, called the critical probability, will be compared against a uniform random number $u_t(i)$ generated from a uniform $[0, 1]$ distribution. The day of interest will be wet if this random number is smaller than the critical probability, and dry otherwise:

$$p_c(i) = \begin{cases} p_{01}(i), & \text{if } x_{t-1}(i) = 0 \\ p_{11}(i), & \text{if } x_{t-1}(i) = 1 \end{cases} \quad (5.1)$$

and

$$x_t(i) = \begin{cases} 1, & \text{if } u_t(i) \leq p_c(i) \\ 0, & \text{otherwise} \end{cases} \quad (5.2)$$

The precipitation amount in a given wet day is derived by inverting a precipitation probability distribution function. In the case of an exponential function, the amount is:

$$r_t(i) = -\ln(1 - v_t(i)) / \lambda_t(i) \quad (5.3)$$

where:

$r_t(i)$ is the synthetic precipitation amount at site i on day t ;

$v_t(i)$ is a uniform $[0, 1]$ random number, independent from $u_t(i)$;

$\lambda_t(i)$ is the rate parameter of the exponential distribution function at site i on day t .

5.2.2 Uni-site generation of daily temperature and solar radiation data

WGEN generates maximum temperature, minimum temperature and solar radiation using the weakly stationary generating process proposed by Matalas (1967):

$$\chi_{p,i}(j) = A \chi_{p,i-1}(j) + B \varepsilon_{p,i}(j) \quad (5.4)$$

where:

$\chi_{p,i}(j)$ and $\chi_{p,i-1}(j)$ are (3×1) matrices of maximum temperature ($j=1$), minimum temperature ($j=2$) and solar radiation ($j=3$) residuals for days i and $i-1$ of year p . The residuals are obtained from:

$$\chi_{p,i}(j) = (X_{p,i}(j) - \bar{X}_i(j)) / \sigma_i(j) \quad (5.5)$$

where:

$X_{p,i}(j)$ is the meteorological variable;

$\bar{X}_i(j)$ is the mean of $X_{p,i}(j)$;

$\sigma_i(j)$ is the standard deviation of $X_{p,i}(j)$.

WGEN computes the residuals for wet and dry days separately, as temperatures and solar radiations are affected by the wet/dry state of the day.

$\varepsilon_{p,i}(j)$ is the (3×1) matrix of independent standard normal random numbers $N[0,1]$ for day i of year p ;

A and B are (3×3) matrices whose elements are defined from lag 0 and lag 1 serial and cross-correlation coefficient matrices of observed residuals :

$$A = M_1 M_0^{-1} \quad (5.6)$$

$$BB^T = M_0 - M_1 M_0^{-1} M_1^T \quad (5.7)$$

where M_0 and M_1 are matrices of lag 0 and lag 1 serial and cross-correlations of observed residuals, respectively (Richardson, 1981).

5.3 Multi-site weather generator (Khalili *et al.* 2007, 2008a)

5.3.1 Multi-site generation of daily precipitation data

Khalili *et al.* (2007) generalized the WGEN presented above using the spatial autocorrelation concept. This statistic has the advantage of describing the spatial dependence observed over an entire watershed with a single number, commonly the Moran's I (Moran, 1950; Odland, 1988; Griffith, 2003), defined as:

$$I = \frac{\sum_{i=1}^n (x_i - \bar{x}) \sum_{j=1}^n w_{ij} (x_j - \bar{x}) / \sum_{i=1}^n \sum_{j=1}^n w_{ij}}{\sum_{i=1}^n (x_i - \bar{x})^2 / n} \quad (5.8)$$

where x_i denotes the observed value of a single variable X at location i , \bar{x} is the average of the x_i over n locations and w_{ij} is the spatial weight between two locations i and j .

Therefore, to compute the spatial autocorrelation between a set of n weather stations, we should select a weight matrix $W(n, n)$ whose elements are w_{ij} . The spatial weights describe the degree of interaction between each pair of stations. These weights may be formulated using the arrangement of the weather stations in space, but other relationships between the stations, such as the precipitation correlations, are recommended by Khalili *et al.* (2007, 2008a). This choice allows the pair station correlations to be preserved in the synthetic time series.

More specifically, the multi-site generation approach of daily occurrence and amount precipitation data uses spatially autocorrelated random numbers, in equations (5.2) and (5.3), whose spatial autocorrelations permit the reproduction of observed daily spatial autocorrelations in the simulated precipitation time series. The monthly interstation correlations are automatically reproduced using monthly weight matrices modelled with monthly interstation correlations of daily precipitation data (Khalili *et al.*, 2007). A moving average process (Cliff and Ord, 1981; Cressie, 1993) is used with suitable coefficients (Khalili *et al.*, 2007) to generate spatially autocorrelated random numbers, such that:

$$V = \gamma \times W \times u + u \quad (5.9)$$

where:

$V(n, 1)$ is a vector of n spatially autocorrelated random numbers to be used for n stations;

$W(n, n)$ is a weight matrix;

$u(n, 1)$ is a vector of n independent and uniformly $[0, 1]$ distributed random numbers.

Because the usual exponential function does not allow the spatial intermittence property for the precipitation amounts to be fulfilled, spatial exponential functions are developed. The rate

parameters of these functions are generated according to the spatial dependence computed for the occurrence values at the set of stations on a given day (Khalili *et al.*, 2008a). This approach was proposed because of the strong dependence existing between the mean precipitation amounts at a given station and the occurrence states at the set of stations. The spatial autocorrelation concept is again used to specify this spatial dependence of the occurrence processes over the watershed.

5.3.2 Multi-site generation of daily temperature and solar radiation data

The multi-site generation approach of daily temperature and solar radiation data (Khalili *et al.*, 2008a) considers the weakly stationary generating process (Eq. (5.4)) as used in the WGEN, but with spatially autocorrelated random numbers in $\varepsilon_{p,i}(j)$. This approach does not involve any change in the model size. Three spatial moving average processes are used to model the maximum temperature, minimum temperature and solar radiation, such that:

$$V_{T_{\max}} = \gamma_{T_{\max}} \times W \times u_{T_{\max}} + u_{T_{\max}} \quad (5.10)$$

$$V_{T_{\min}} = \gamma_{T_{\min}} \times W \times u_{T_{\min}} + u_{T_{\min}} \quad (5.11)$$

$$V_{Sr} = \gamma_{Sr} \times W \times u_{Sr} + u_{Sr} \quad (5.12)$$

where :

$V_{T_{\max}}(n,1)$, $V_{T_{\min}}(n,1)$ and $V_{Sr}(n,1)$ are vectors of n spatially autocorrelated random numbers required to generate the maximum temperature, minimum temperature and solar radiation variables, respectively, at the n stations;

$W(n,n)$ is a weight matrix;

$\gamma_{T_{\max}}$, $\gamma_{T_{\min}}$ and γ_{Sr} are the moving average coefficients to be used for maximum temperature, minimum temperature and solar radiation variables, respectively;

$u_{T_{\max}}(n,1)$, $u_{T_{\min}}(n,1)$ and $u_{Sr}(n,1)$ are vectors of n independent and normally distributed random numbers to be used for maximum temperature, minimum temperature and solar radiation variables, respectively.

The regionalization is carried out using equation (5.4) along with $\varepsilon_{p,i}(j)$ containing 3 random numbers, from $V_{T_{\max}}(n,1)$, $V_{T_{\min}}(n,1)$ and $V_{S_r}(n,1)$, respectively. For a station S , for example, the multi-site model is expressed by:

$$\chi_{p,i}(j) = A \chi_{p,i-1}(j) + B \begin{pmatrix} V_{T_{\max}}(S) \\ V_{T_{\min}}(S) \\ V_{S_r}(S) \end{pmatrix} \quad (5.13)$$

The 3 elements, $V_{T_{\max}}(S)$, $V_{T_{\min}}(S)$ and $V_{S_r}(S)$, used in the $\varepsilon_{p,i}(j)$ vector, for a station S , are independent, but each is spatially autocorrelated with the remaining $n - 1$ components of the same vector type, $V_{T_{\max}}$, $V_{T_{\min}}$ or V_{S_r} respectively, which will be used for the remaining stations. Using appropriate moving average coefficients (Khalili *et al.*, 2008a), the spatial autocorrelations of these numbers allow the generation of synthetic weather data with daily spatial autocorrelations akin to those observed. The interstation correlations are automatically reproduced using monthly weight matrices expressed as monthly interstation correlations of daily weather variables.

5.4 Hydrological modeling

5.4.1 Description of the study area

The multi-site generation approach as described in the previous section has been applied to the Peribonca River Basin (26 000 km²). The watershed is located north of the Lac-Saint-Jean, in the Canadian province of Quebec, and has four sub-basins: Passes-Dangereuses (11 000 km²), Lac Manouane (5 000 km²), Chute-du-Diable (9 700 km²) and Chute à la Savane (1 300 km²). The 7 weather stations used in this study are located either within or around the Chute-du-Diable sub-basin. This watershed was therefore used for the hydrological modelling experiment (Figure 5.1).

The Chute-du-Diable watershed has a cold and snowy climate, with monthly average precipitation amounts ranging from 38 to 142 mm. Its soil is typically composed of till and sand (Table 5.1). The Chute-du-Diable land use is dominated by the evergreen class, followed by the mixed class (Table 5.2).

5.4.2 Hydrotel model application

The mainly reason for selecting the Hydrotel model (Fortin *et al.*, 2001a, 2001b) for this study is its ability to take into account the spatial variability of meteorological and hydrological processes. The distributed structure of this model allows a more thorough investigation of the performance of the multi-site weather generator. Another advantage of this model is its ability to be adapted to available data and to offer a variety of physical, conceptual and empirical hydrological processes.

Applying the Hydrotel model previously required the application of Physitel, in order to prepare the Chute-du-Diable watershed physiographic database. This software program uses a digital elevation model (DEM) to compute the slope of all the constituted cells and then to determine the flow directions through them. Physitel used these flow directions and the location of the watershed outlet to determine the complete structure of the Chute du Diable watershed. 310 Relatively Homogenous Hydrological Units (RHHU) were then constituted in this watershed. These very small sub-watersheds permit an adequate simulation and mapping of the aerial distribution of hydrological processes (Fortin *et al.*, 2001a). The observed streamflows of this watershed are proposed by ALCAN, which owns and operates the meteorological and streamflow stations. Figure 5.2 illustrates the topography of the Chute-du-Diable watershed and its river reaches, as obtained from Physitel.

In addition to the watershed database provided by Physitel, Hydrotel requires hydrometeorological data. Precipitation and temperature data from the selected stations are used as well as the streamflows in the Chute-du-Diable watershed. In fact, the model calibration is realized using the observed hydrometeorological data.

Hydrotel submodels can be selected according to their ability to simulate the hydrological processes and also to available data. First, the interpolation of the meteorological data to each RHHU is required in order to ensure the spatial distribution of meteorological data on the watershed, and to that end, either Thiessen polygons or weighted means from the nearest three stations can be used. Hydrotel then separates the precipitation data at each time step into rainfall and snowfall according to the maximum and minimum temperature data at that time. Snow accumulation and melt are computed over each RHHU using a mixed degree-day-energy-budget approach. Depending on data availability, Hydrotel offers various options to simulate the potential evapotranspiration for each RHHU, such as Thornthwaite, Hydro-Québec, Linacre, Penman-Monteith and Priestley-Taylor equations. A three-layer vertical water budget submodel (BV3C) simulates the moisture in the soil column at each RHHU. The BV3C submodel computes the water amounts available for infiltration, runoff, interflow and base flow. Surface and subsurface flows are propagated from one RHHU to the next and toward the hydrographic network using the kinematic wave equation. The flow in the hydrographic network is routed to the watershed outlet using the diffusive wave equation.

5.5 Results and discussion

The application of the Hydrotel model in this paper was carried out in three steps. First, the model was calibrated for the Chute-du-Diable watershed using observed meteorological and streamflow data from 1963 to 1976. This particular period was selected because it was also used for the multi-site weather generation process. More specifically, the spatial autocorrelations of both precipitation and temperature data were computed using a shared recorded period between the selected weather stations, which ran from 1963 to 1976. The Nash-Sutcliffe coefficient was equal to or higher than 0.8 for the full calibration period, and for each year separately. Validation was performed using the 1980-1998 hydrological years, with an overall Nash-Sutcliffe coefficient of 0.77.

Once calibrated, Hydrotel was then driven with weather data produced by the multi-site and uni-site weather generators, as step two and three of the modelling experiment. The resulting

streamflows were then compared to those simulated using the observed weather data, referred to as the “reconstituted” streamflows, in order to verify the performance of each of the weather generation approaches in hydrological modelling.

A visual assessment of the two simulation results using multi-site and uni-site generated weather data revealed obvious impacts consecutive to using the uni-site weather generator. In fact, using weather data produced by WGEN leads to a significant underestimation of the summer-autumn peak flows (see Figure 5.3). On the other hand, using weather data produced by the multi-site weather generator resulted in simulated summer-autumn flows, which better resembled the reconstituted flows. Note that it is expected that the curves should not exactly match because of the stochastic process driving the weather generators. The reconstituted streamflows are those obtained for the year 1971.

The behaviour of the hydrograph simulated using uni-site generated weather data can be explained by the very nature of the uni-site weather generator, which does not take into account the spatial dependence existing between the weather stations, particularly for the precipitation processes, which exhibit a high spatial variability, and are the main driving factor in the hydrological modelling. The weather data are generated at a single site independently of the others, while weather events typically exhibit a spatially organised pattern over a region. The surrounding weather stations should be interdependent because they experience the same or very nearly the same weather events (Odland, 1988).

Monitoring precipitations over the watershed at each step of the simulation run was used to look for closer simulation details. For instance, the highest total amounts of precipitations over the entire watershed recorded in September as rainfall was retrieved from the observed and the simulated data. The heavy total rainfall amounts over the entire watershed was much more significant for observed (290 mm) and multi-site simulated patterns (229 mm) than for uni-site simulated ones (94 mm). Figures 5.4a, 5.4b and 5.4c display the spatial distribution of these heavy total rainfall amounts for observed, multi-site and uni-site simulated cases, respectively, over the Chute-du-Diable watershed. Clearly, the observed precipitations

exhibit a significant spatial dependence, and the multi-site weather generator succeeded in simulating the precipitation processes with a realistic spatial distribution, while the uni-site weather generator resulted in a local rain, mainly in the south-east of the watershed. Therefore, ignoring the spatial dependence over the Chute-du-Diable watershed will affect the magnitude of the simulated flow.

Further analyses were performed to evaluate the hydrological response of the Hydrotel model to the observed, multi-site and uni-site generated weather data. Although the weather generators can produce time series of infinite length, 50 years of data were generated, from which streamflows were simulated using Hydrotel. These flows were statistically compared to 30 year time series of reconstituted flows, spanning from 1960 to 1989, in order to allow a more “objective” evaluation of the performance of the weather generators.

Flood frequency analysis, which may be the better indicator of extreme values, was carried out in order to compare the likelihood of extreme streamflow events. The Pearson III frequency distribution was found to fit better with the summer and spring data while the Gumbel frequency distribution was retained for autumn. The method of moments was used for parameter estimation, and Weibull or Hazen equations were used for the plotting position. Distributions were fitted on the normal probability paper. Summer streamflow frequency results are presented in Figure 5.5. Figures 5.5a and 5.5b show a slight difference between the streamflow distribution slopes using observed and multi-site generated weather data, while a smoother slope is obtained with uni-site generated weather data (Figure 5.5c). Similar results are also observed for the autumn season (Figure 5.6).

Tables 5.3 and 5.4 show that extreme streamflows simulated using the uni-site weather generator are considerably lower than those simulated with observed weather data for both the summer and autumn seasons and for all the return periods. The underestimation is less significant using the multi-site weather generator, especially in summer.

The underestimation of extreme streamflows obtained mainly in autumn, using the multi-site weather generator, was expected because of its limited ability to simulate the more extreme daily precipitation amounts using the exponential distribution function. This underestimation is much more pronounced with the uni-site generated input because of the independence of the weather stations, which can result in a large rainfall at a given station, but none at all at another station nearby, on a same day. However, the marked difference in autumn is caused by the large reconstituted streamflow depths (998 m³/s and 981 m³/s) generated on November 5th, 1966, and on October 20th, 1967, respectively. These two outliers (Figure 5.6a), which are very distant from the remaining data, can more probably be due to a misrepresentation of the weather stations. When these outliers were omitted, similar frequency results were obtained with both reconstituted and simulated streamflows using the multi-site weather generator (Table 5.4).

Despite the slight underestimation of maximum streamflows using the multi-site weather generator, the results clearly confirm the importance of considering the spatial dependence of the weather stations in the flow modelling process. Major errors can result from using the uni-site weather generator. It was shown that extreme streamflows from lower to upper return periods are poorly reproduced by this generator.

Regarding the spring season, a good agreement was obtained between the reconstituted and simulated streamflows using the multi-site weather generator. With the uni-site generated weather data, extreme streamflows decreased as the return period increased (see Table 5.5 and Figure 5.7).

Hydrotel responses to observed and simulated weather data was also evaluated using streamflow averages. Figure 5.8 illustrates the monthly averages of daily streamflows obtained using observed as well as multi-site and uni-site generated data. The three kinds of streamflows are generally in good agreement. The monthly averages of simulated streamflows obtained using the multi-site weather generator are the closest to those reconstituted. However, an overestimation of the streamflow average is obtained for May,

and is more pronounced for the simulated streamflows obtained with the uni-site weather generator.

This overestimation does not mean that multi-site and uni-site weather generators inflate the volume in the Chute-du-Diable watershed, but rather, it is related to the limited variability of the generated air temperature data. The winter and spring streamflows are mainly linked to the accumulation and melt of snow cover. This melt is modelled by Hydrotel using air temperature as the driving meteorological variable. Further investigations revealed that the simulated air temperature data did not display as much variability as the observed records, particularly with the uni-site weather generator. This limitation also affects the streamflow variability, as shown by Figure 5.9, which is mostly underestimated using the uni-site weather generator.

Figure 5.10 shows the standard deviations of observed and simulated daily air temperatures. These data are derived from Hydrotel as the air temperatures over the entire watershed. It therefore appears that observed and multi-site generated air temperatures show considerable variability, and are generally in good accordance, except for December, January, February and March, which constituted the winter season for this watershed. Uni-site generated air temperatures show lower variability.

The misreproduction of the variability obtained with the uni-site weather generator and slightly in winter, with the multi-site generator, affects the melt of the snow cover, and consequently, the timing of the spring peak flow. In fact, using the weather generated data, the spring peak flow for each simulated year occurs almost always in May, but from April to June, using observed data because of the mid-winter thaw. Therefore, using the weather generators, the occurrence of the spring peak flows annually in May causes the highest streamflow average at this time, as shown in Figure 5.8. However, from observed data, the much longer snowmelt flood occurrence period causes the highest streamflows to be spread out between April, May and June in the observed years, causing a moderate streamflow average peak.

Figure 5.10 presents another advantage of the multi-site weather generator. It appears that taking the spatial dependence between the synthetic temperature data into account satisfactorily reproduces the air temperature variability. The latter may be improved to be adequately reproduced in winter to allow a reasonable timing of spring peak flows and more accurate streamflow standard deviations.

Finally, it is important to note the substantial errors, which can be made using uni-site weather generators for hydrological modelling. The importance of multi-site generated weather data is thus recognized, and the multi-site weather generator should be used when considering hydrological modelling in climate change, hydraulic design, water resource analysis and other meteorology dependent projects.

5.6 Conclusion

This paper presents a comparative study to assess the performance of multi-site and uni-site weather generators for streamflow simulation. The uni-site weather generator used is the WGEN (Richardson, 1981), while the multi-site generator is the one developed by Khalili *et al.* (2007, 2008a). This multi-site weather generator is also based on the WGEN approach, but uses spatially autocorrelated random numbers to reproduce the spatial autocorrelations computed over the watershed as well as the monthly pairwise correlations.

The Chute-du-Diable watershed (9 700 km²) located in the Canadian province of Quebec was used for this study. The Hydrotel model was selected for hydrological modelling. The distributed nature of this model provides a more in-depth understanding of the effect of spatially dependent meteorological data on the runoff generating mechanism. To carry out the comparison, 30 years of observed meteorological data were used. The response of the Hydrotel model to these observed data provided reconstituted streamflows against which simulated flows using generated meteorological data were compared. Weather generators can be used to simulate records of any desired length, but only the results for 50 simulated years were presented here.

The results of this study indicate an underestimation of summer-autumn peak flows using the uni-site weather generator. However, realistic summer-autumn simulated peak flows were obtained with the multi-site synthetic weather data since the spatial autocorrelations over the watershed and the correlations between each pair of stations were adequately reproduced by the multi-site weather generator.

Frequency analyses indicated a strong underestimation of extreme streamflows in summer and autumn using the uni-site weather generator. The summer floods were however reasonably reproduced using the multi-site weather generator. A slight underestimation was obtained for the autumn season, mainly due to the exponential function used to model the precipitation amounts. The magnitude of the spring floods was adequately simulated using the multi-site weather generator while an underestimation still resulted from the use of the uni-site generator.

Monthly streamflow averages were computed for the reconstituted streamflows as well as for the simulated ones using the multi-site and uni-site weather generators. Generally, there was good agreement between the observed and the simulated streamflow averages. An overestimation of the simulated streamflow averages was obtained in May, and was more pronounced using the uni-site weather generator. This overestimation was explained by a limited variability of the simulated air temperature data, particularly with the uni-site weather generator, which directly affected the timing of the spring freshet.

This study demonstrated the superiority of the multi-site weather generation approach over the uni-site approach in hydrological modelling. Ignoring the spatial dependence between the weather locations yields drastic errors in any work using hydrological modelling, and it is therefore strongly recommended that any further study, especially for the assessment of climate change impacts on hydrology regimes, should use the multi-site weather generator.

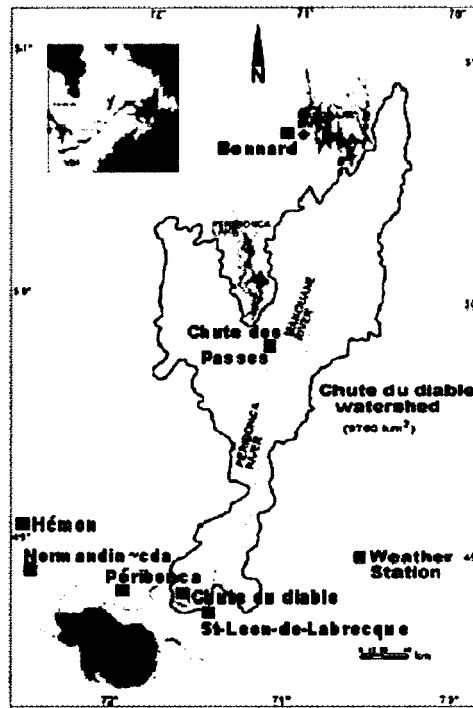


Figure 5.1 *Chute du Diable watershed with locations of weather stations.*

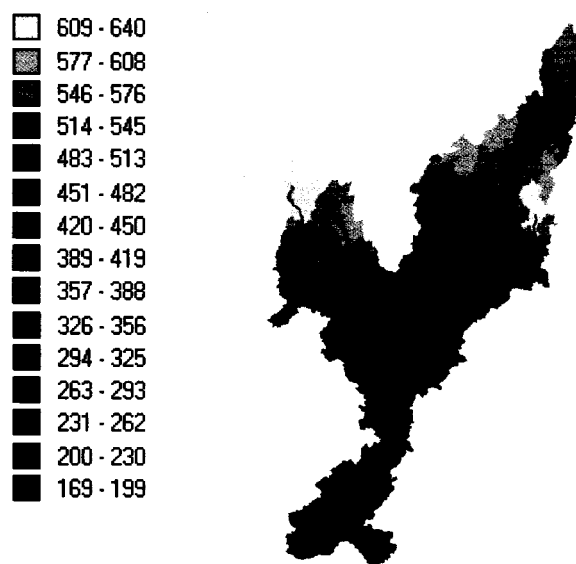


Figure 5.2 *River reaches and altitudes (m) of the Chute du Diable watershed obtained from Physitel.*

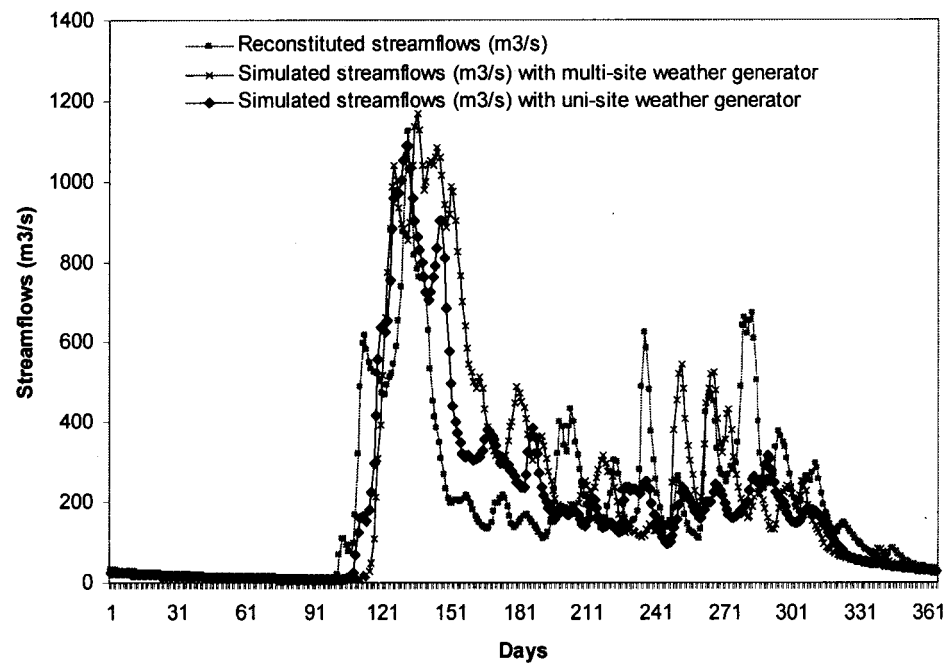


Figure 5.3 *Reconstituted and simulated streamflows using multi-site and uni-site weather generators.*

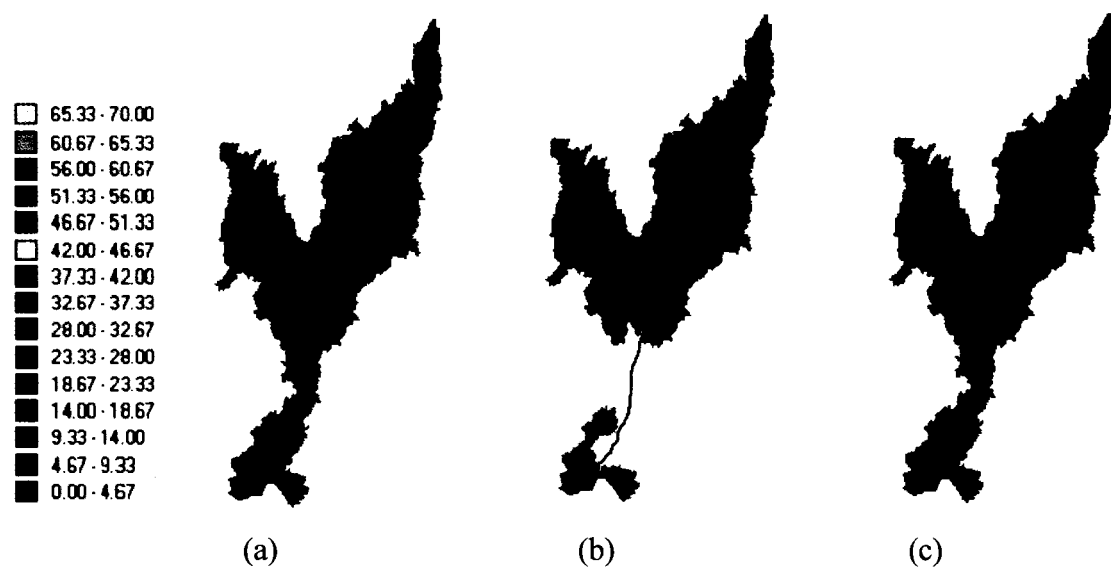


Figure 5.4 *Spatial distribution of heavy total rainfall amounts occurring in September over the Chute du Diable watershed from (a) observed data, (b) multi-site generated data, and (c) uni-site generated data.*

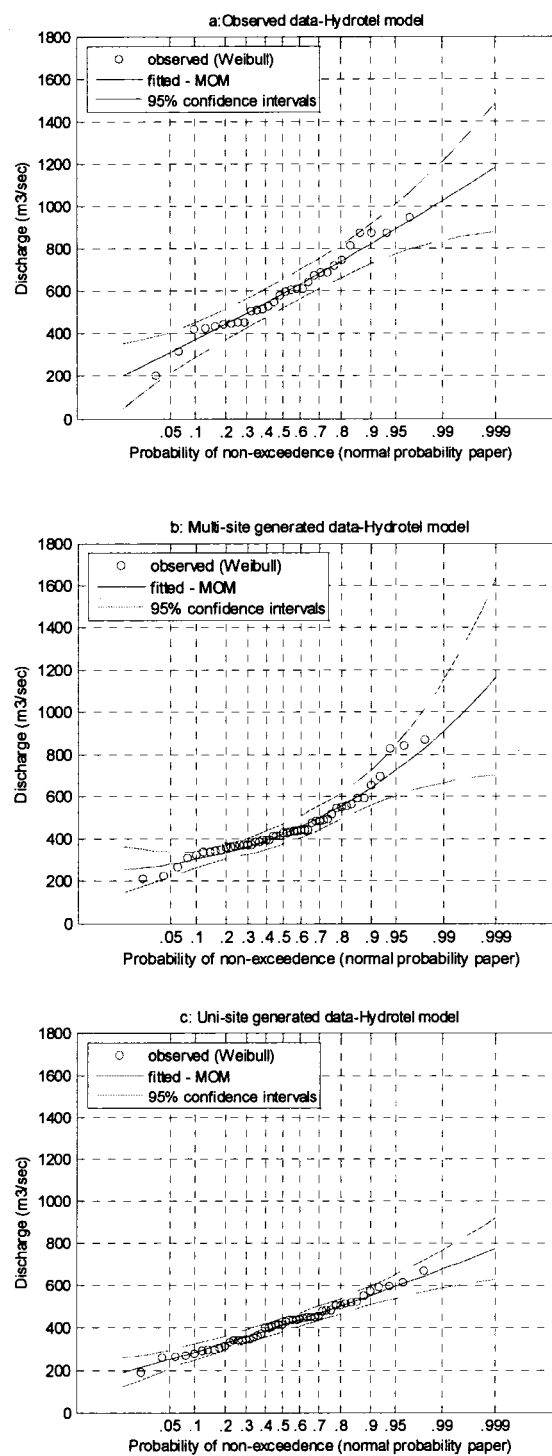


Figure 5.5 Summer streamflow frequency distributions using (a) observed data, (b) multi-site generated data, and (c) uni-site generated data.

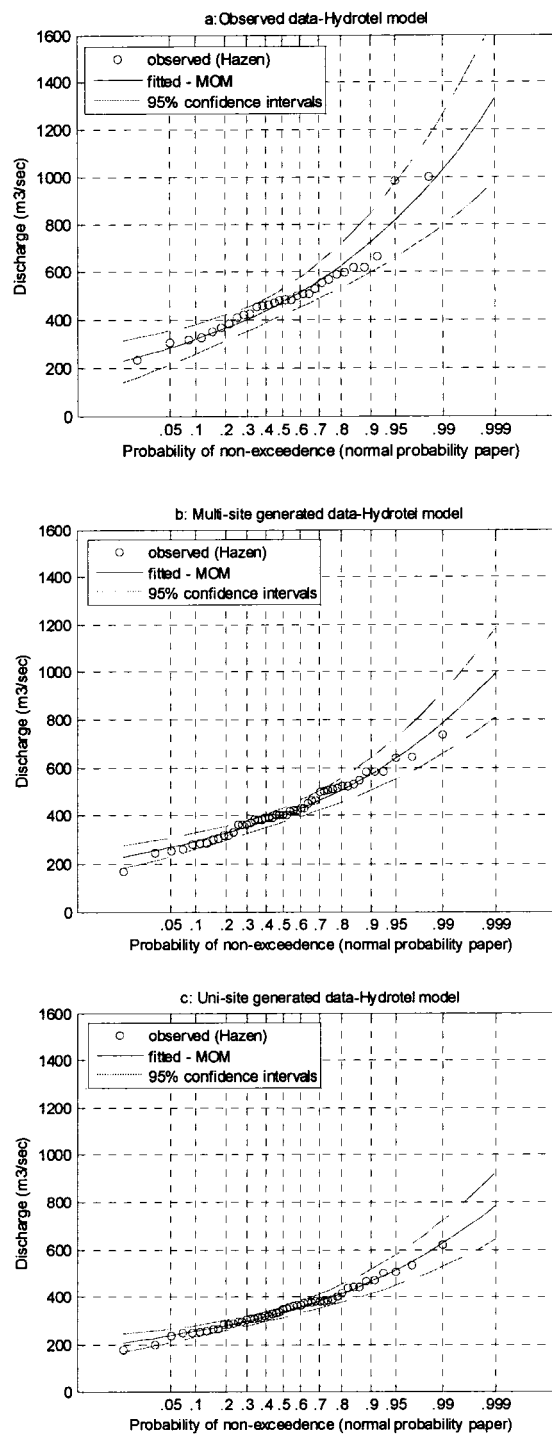


Figure 5.6 *Autumn streamflow frequency distributions using (a) observed data, (b) multi-site generated data, and (c) uni-site generated data.*

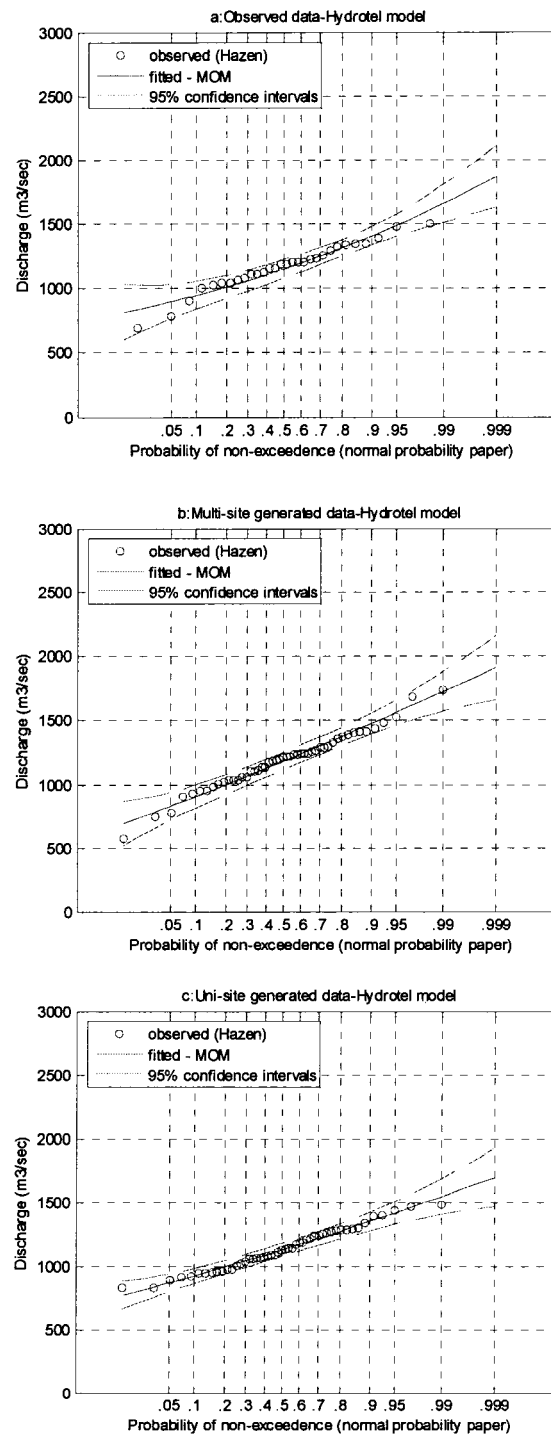


Figure 5.7 *Spring streamflow frequency distributions using (a) observed data, (b) multi-site generated data, and (c) uni-site generated data.*

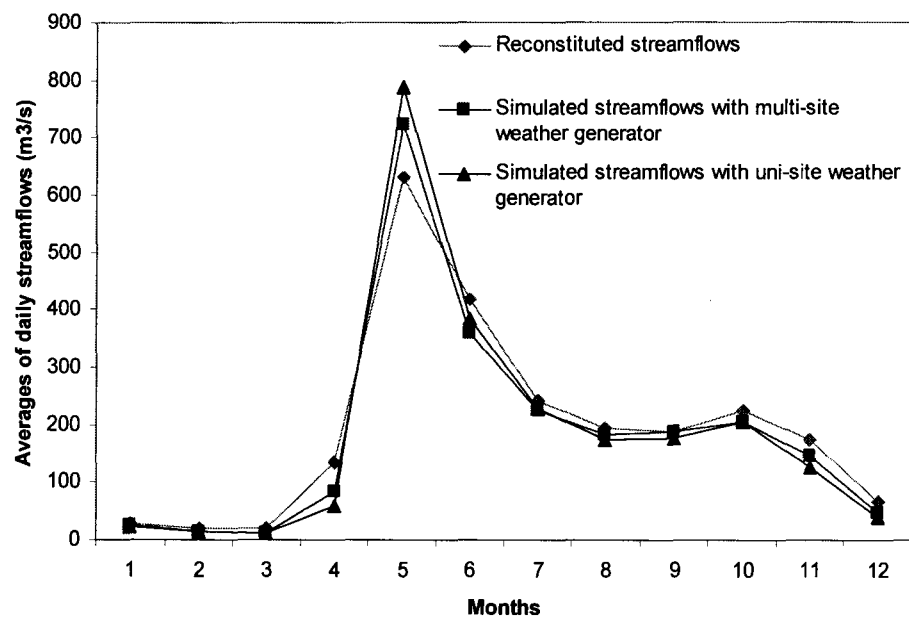


Figure 5.8 *Monthly streamflow averages.*

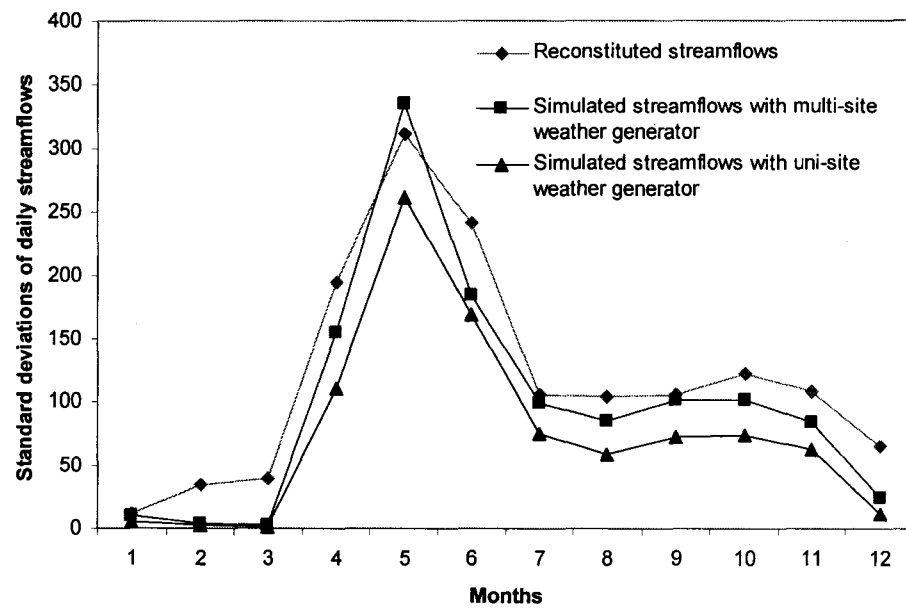


Figure 5.9 *Monthly streamflow standard deviations.*

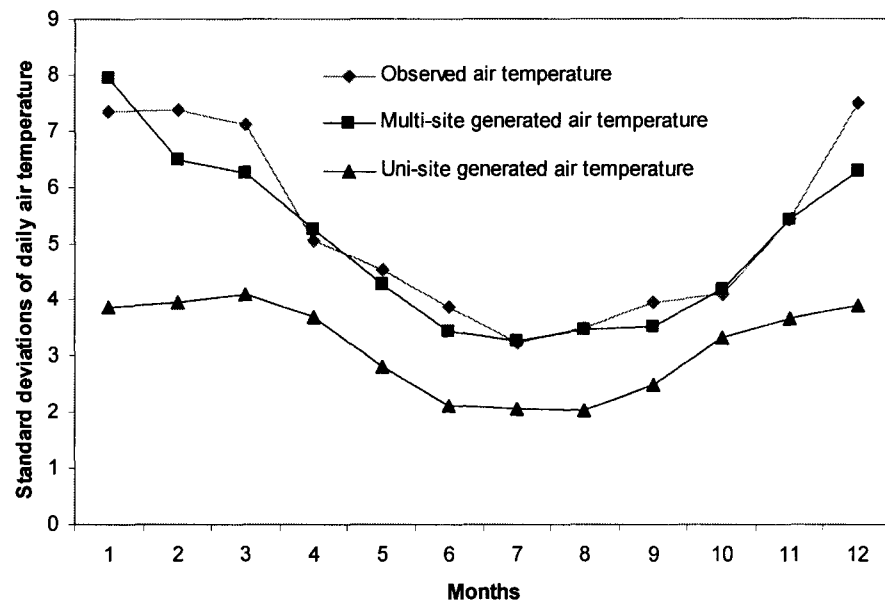


Figure 5.10 *Monthly air temperature standard deviations.*

Tableau 5.1

Soil type on Chute du Diable Watershed

Soil type	Area (%)
Till_sand	94.5
Till and organic deposit	4.5
Organic deposit	0.3
Till_sand_gravel	0.6

Tableau 5.2

Land-use occupation on Chute du Diable Watershed

Land-use class	Area (%)
Crop_Grass	0.1
Crop_Wood	2.4
Shrubland	0.8
Deciduous	2.2
Evergreen	65
Mixed	26.1
Water	3.4

Tableau 5.3

Summer flood frequency using observed data,
multi-site, and uni-site generated data in Hydrotel model

Return period	Summer		
	Observed data	Multi-site generated data	Uni-site generated data
1	308.46	279.28	252.25
2	585.97	422.59	412.82
5	738.13	553.03	503.05
10	820.49	640.19	552.46
20	889.97	723.04	594.46
50	969.80	828.88	643.03
100	1023.98	907.28	676.19
200	1074.24	984.78	707.09
500	1136.02	1086.38	745.25

Tableau 5.4

Autumn flood frequency using observed data with outliers (30 pieces of data) and without outliers (28 pieces of data), multi-site, and uni-site generated data in Hydrotel model

Return period	Autumn			
	30 Observed data	28 Observed data	Multi-site generated data	Uni-site generated data
1	283.08	328.37	268.22	236.66
2	474.31	449.56	401.17	336.58
5	622.38	543.39	504.12	413.95
10	720.42	605.52	572.28	465.17
20	814.46	665.12	637.66	514.31
50	936.19	742.26	722.29	577.91
100	1027.40	800.06	785.71	625.57
200	1118.29	857.66	848.90	673.06
500	1238.19	933.64	932.26	735.71

Tableau 5.5

Spring flood frequency using observed data,
multi-site, and uni-site generated data
in Hydrotel model

Return period	Spring		
	Observed data	Multi-site generated data	Uni-site generated data
1	896.13	835.39	873.82
2	1146.45	1183.90	1132.26
5	1308.70	1372.24	1275.93
10	1403.59	1473.45	1354.22
20	1487.41	1558.47	1420.54
50	1587.87	1655.74	1497.02
100	1658.54	1721.53	1549.10
200	1725.87	1782.39	1597.54
500	1810.95	1857.00	1657.24

CONCLUSION

Les générateurs de climat sont des outils statistiques qui permettent de fournir de longues séries temporelles de données climatiques, souvent rarement ou nullement disponibles, qui décrivent aussi bien le climat présent que le climat futur. Ce dernier pourra être simulé si les générateurs de climat ont tenu compte des scénarios de changements climatiques fournis par les modèles de circulation générale. La génération des données climatiques pour des études hydrologiques, agraires ou environnementales ne doit pas être prise à la légère. Des projets aussi bien décisionnels que coûteux doivent être étudiés avec des bases de données qui assurent une meilleure représentation des phénomènes météorologiques qui, par exemple, influencent ou influenceront le comportement hydrologique des bassins versants. Cette qualité ne peut être offerte par les générateurs de climat uni-site pour les bassins versants de moyenne et grande taille, ce qui entrave la représentation réaliste des processus hydrologiques résultants. La négligence de la cohérence spatiale entre les données climatiques générées se répercute aussi bien sur la variabilité spatiale des événements climatiques que sur l'ampleur et la fréquence des débits extrêmes, comme il a été démontré dans cette thèse.

L'objectif de cette thèse était de développer une approche de génération multi-site des données climatiques. Les critères recherchés pour l'approche développée, après une littérature exhaustive des différents travaux réalisés, étaient la performance, la simplicité d'implantation et la parcimonie.

Cet objectif a été réalisé en utilisant le concept d'autocorrélation spatiale. L'autocorrélation, comme son nom l'indique, est la corrélation entre les valeurs d'une même variable. Elle peut être temporelle si ces valeurs sont attribuées au temps ou bien spatiale si elles sont attribuées à l'espace. L'autocorrélation spatiale représente donc le degré de dépendance des valeurs, d'une même variable, en tenant compte de leurs positions dans l'espace. Elle peut être mesurée par des indicateurs statistiques comme l'indice de Moran.

Avec leur caractère géographique, les données climatiques sont forcément caractérisées par l'autocorrélation spatiale induite par les phénomènes climatiques qui se développent au dessus d'une région au lieu d'un site en particulier. L'indice de Moran est donc employé, dans cette thèse, pour mesurer l'autocorrélation spatiale des variables climatiques représentées par les occurrences de précipitations, les quantités de précipitations, les températures maximales, les températures minimales et les radiations solaires.

La vision principale de la régionalisation du générateur de climat uni-site était de générer des données climatiques en préservant les autocorrélations spatiales observées sur le bassin versant en étude. Cette régionalisation est réalisée en utilisant le générateur de climat uni-site mais avec des nombres aléatoires spatialement autocorrélés pour reproduire les autocorrélations spatiales des processus météorologiques observés. Le processus de moyenne mobile spatialisé est utilisé pour la génération des nombres aléatoires spatialement autocorrélés. Des relations ont été trouvées entre les coefficients de ce processus et les autocorrélations spatiales des variables météorologiques simulées, ce qui a permis d'identifier les coefficients adéquats qui vont reproduire les autocorrélations spatiales observées de ces variables météorologiques.

Le privilège accordé à l'autocorrélation spatiale est qu'elle représente, à l'aide d'un seul nombre, la dépendance spatiale sur l'étendue du bassin versant. Sa mesure requiert, toutefois, la détermination d'une matrice, dite de poids ou de voisinage, faisant intervenir l'interrelation entre chaque couple de stations. Tel que démontré dans cette thèse, nous avons bénéficié de cette structure spatiale pour générer des données climatiques en préservant aussi bien les autocorrélations spatiales observées sur le bassin versant que les corrélations entre chaque couple de stations météorologiques. Ces corrélations sont reproduites en utilisant des matrices de poids mensuelles basées sur les corrélations mensuelles observées entre les processus météorologiques.

L'application de cette méthode de régionalisation est effectuée pour le bassin versant de la rivière Péribonca situé dans la province du Québec, Canada. Sept stations météorologiques

ont été sélectionnées sur ce bassin avec des enregistrements observés variant de 17 à 57 ans. Le calcul des autocorrélations spatiales observées a été effectué sur une période commune pour les sept stations, soit de 1963 à 1976.

L'approche de génération multi-site est réalisée avec succès pour les occurrences et les quantités de précipitations (article 1 : Khalili *et al.* 2007). Les autocorrélations spatiales journalières et mensuelles sont bien reproduites ainsi que les corrélations mensuelles entre chaque couple de stations. Une bonne concordance est obtenue entre les statistiques des séries météorologiques observées et simulées comme les nombres mensuels de jours pluvieux, les quantités de précipitations totales mensuelles, les variances des précipitations journalières et les probabilités jointes des processus d'occurrences des précipitations. L'intermittence spatiale des quantités de précipitations n'est toutefois pas reproduite avec la fonction exponentielle simple, des fonctions exponentielles spatiales mensuelles sont développées pour chaque station en tenant compte des autocorrélations spatiales, des occurrences de précipitations, calculées sur tout le bassin. Cette méthode s'est révélée prometteuse puisque la moyenne des précipitations, à un site donné, dépend des états d'occurrences sur les sites avoisinants. Des résultats satisfaisant sont obtenus par cette méthode (première partie de l'article 2 : Khalili *et al.* 2008a).

Pour les autres variables climatiques, alors que la méthode de génération multi-site paraissait une tâche ardue, l'approche de régionalisation développée dans cette thèse l'a considérablement facilitée en utilisant l'autocorrélation spatiale (deuxième partie de l'article 2 : Khalili *et al.* 2008a). En effet, le même modèle autorégressif, proposé par Richardson (1981), est utilisé mais avec des nombres aléatoires spatialement autocorrélés, permettant de reproduire les autocorrélations spatiales observées dans les séries de données synthétiques. Aucune augmentation de la taille du modèle autorégressif n'est requise même si d'autres éventuelles variables climatiques doivent être prises en compte ou bien d'autres stations sont ajoutées dans le réseau. L'utilisation des matrices de poids mensuelles, exprimées par les corrélations mensuelles observées des températures, a permis de reproduire également les corrélations mensuelles entre chaque couple de stations.

Il importe de souligner que cette approche de régionalisation, basée uniquement sur le concept d'autocorrélation spatiale, n'a requis que la détermination des coefficients du processus de moyenne mobile spatialisé qui vont reproduire les autocorrélations spatiales observées entre les processus météorologiques sur l'ensemble des stations. Ces coefficients appropriés sont obtenus avec des relations mensuelles entre les valeurs possibles de ces coefficients et les autocorrélations spatiales des processus météorologiques.

Le générateur de climat multi-site développé ici a aussi indirectement été validé en le comparant au générateur uni-site dans le cadre de la modélisation hydrologique. Pour ce faire, le modèle hydrologique distribué, Hydrotel (Fortin *et al.*, 2001a, 2001b), a été calibré pour le sous bassin de la Chute-du-Diable duquel les stations utilisées sont proches comparativement aux autres sous-bassins du bassin versant de la rivière Péribonca. La structure distribuée du modèle Hydrotel a permis une comparaison plus exhaustive entre les données météorologiques synthétiques issues des deux générateurs, multi-site et uni-site, dans le sens où elle a fait surgir l'effet de la dépendance spatiale. Le calage et la validation du modèle Hydrotel ont été réalisés avec les données météorologiques observées aux sept stations et les données hydrométriques observées au sous-bassin. Les débits obtenus lors de la phase de calibration, appelés débits reconstitués, ont été considérés comme référence pour cette étude comparative. La version calibrée du modèle Hydrotel a ensuite été utilisée avec les données météorologiques simulées par les générateurs multi-site et uni-site.

Les résultats ont montré visuellement une sous-estimation évidente des pointes de crues été-automne avec le générateur uni-site. Ce résultat est dû à la négligence de la cohérence spatiale par le générateur uni-site. Cette négligence influence la variabilité spatiale des données climatiques, notamment les précipitations, et par conséquent minimise les débits maximums. Par contre, les débits de pointes obtenus, en utilisant les données climatiques simulées par le générateur multi-site, sont raisonnablement reproduits durant cette période de l'année.

L'analyse de fréquence a montré une forte sous-estimation des débits extrêmes pour les saisons été et automne et pour toutes les périodes de retour, en utilisant les données climatiques simulées par le générateur uni-site. Avec les données du générateur multi-site, une bonne reproduction des débits extrêmes est obtenue en été avec une légère sous-estimation en automne. Cette sous-estimation est due à la fonction exponentielle qui sous-estime les précipitations extrêmes, surtout pour cette saison. Les débits de pointes printaniers sont adéquatement reproduits par les données simulées avec le générateur multi-site. Ils sont, par contre, sous-estimés en utilisant le générateur uni-site.

Quoiqu'ils ne représentent pas un fort critère de performance pour un tel projet qui met l'emphasis sur les crues extrêmes, les débits moyens mensuels obtenus avec les données des générateurs multi-site et uni-site sont également comparés à ceux reconstitués. Une bonne concordance est obtenue entre ces débits moyens. Une surestimation est néanmoins observée au mois de mai, notamment, pour les débits obtenus avec les données du générateur uni-site. Cette surestimation est moins prononcée avec les données du générateur multi-site. Alors qu'elle apparaisse comme une inflation du volume d'eau sur le bassin de la Chute-du-Diable, cette surestimation est plutôt due à la variabilité des températures de l'air. Cette variabilité, profusément sous-estimée par le générateur uni-site, engendre des occurrences de débits de pointes printaniers, pratiquement, au mois de mai à chaque année simulée. Le débit moyen en ce mois de l'année est donc amplifié au détriment des mois d'avril et juin où les crues printanières observées peuvent également survenir. Cette sous-estimation de la variabilité des températures générées uni-site induit également une sous-estimation des variations des débits. La variabilité des températures de l'air est, cependant, bien reproduite par le générateur multi-site, sauf pour la saison d'hiver où une légère sous-estimation a surgi, ce qui explique la légère surestimation du débit moyen du mois de mai et sous-estimation des variations des débits.

Finalement, les résultats de la modélisation hydrologique montrent la performance du générateur de climat multi-site et confirment clairement son utilité pour évaluer avec plus de précision et de fiabilité les fréquences et magnitudes des crues extrêmes. Des erreurs

importantes peuvent ainsi être commises en utilisant des données climatiques spatialement indépendantes, aussi bien pour l'évaluation des impacts des changements climatiques futurs que pour la gestion des ressources hydriques sous les conditions météorologiques actuelles. À la connaissance de l'auteure, cette étude est la première à comparer deux générateurs, multi-site et uni-site, en utilisant un modèle hydrologique distribué.

L'originalité majeure de cette thèse réside dans l'approche de régionalisation, basée sur le concept d'autocorrélation spatiale, appliquée aux données de précipitations (chapitre 3), de températures maximales, minimales et de radiations solaires (chapitre 4). Simple et robuste, cette approche parcimonieuse reste facilement applicable quelque soit la dimension du réseau de stations météorologiques et peut aisément intégrer d'autres éventuelles variables climatiques. Considérée comme première loi de la géographie annoncée par Tobler (1970) «everything is related to everything else, but near things are more related than distant things », l'autocorrélation spatiale des données climatiques est une réalité qu'on ne peut point escamoter ou ignorer.

RECOMMANDATIONS

Les recommandations pour les futures recherches portent sur les points suivants :

- Détermination automatique optimale de la matrice de poids. En effet, d'après les études effectuées pour cette thèse, l'utilisation de matrices de poids exprimées par les corrélations mensuelles observées entre les stations météorologiques a permis de reproduire adéquatement les corrélations observées entre chaque couple de stations. Ces matrices ont été portées à la puissance 2 ou 3 pour optimiser les résultats. Une optimisation automatique à l'aide d'un algorithme est ainsi recommandée. Les matrices de poids peuvent aussi être représentées différemment en utilisant d'autres hypothèses reliant les stations météorologiques.
- Application de la méthode sur d'autres bassins versants avec un grand nombre de stations météorologiques. L'approche de génération multi-site développée dans cette thèse se base sur l'autocorrélation spatiale qui permettra de décrire, avec un seul chiffre, la dépendance spatiale sur l'étendu du bassin versant. Cette propriété permettra ainsi, facilement, d'appliquer l'approche de génération multi-site sur des bassins versants de grande dimension voire sur plusieurs régions à la fois.
- Utilisation d'un modèle de chaîne de Markov avec un ordre plus grand que 1. Il s'agit d'adapter la méthode de régionalisation, développée dans cette thèse, au générateur WeaGETS (Caron, 2006). Aucune modification du principe de régionalisation n'est requise. Il faut, toutefois, savoir que l'ordre optimal de la chaîne de Markov n'est pas nécessairement plus grand que 1. De nombreuses études ont démontré la performance du premier ordre. Le choix de l'ordre est plutôt relié au temps et à l'espace d'étude. Il faut aussi penser à la parcimonie du modèle.
- Utilisation d'une fonction de distribution plus appropriée pour améliorer la génération des quantités extrêmes de précipitations.

- Amélioration de la variabilité des températures, en hiver, pour une meilleure reproduction du temps d'occurrences des débits de pointes printaniers. Cela peut être réalisé en faisant varier les matrices de corrélations de lag 0 et 1 sur toute l'année au lieu d'utiliser des matrices constantes.
- Tenir compte des types de circulation atmosphérique dans la détermination des paramètres du générateur de climat multi-site pour la mise à l'échelle des résultats des MCG et pour refléter le caractère non stationnaire des séries climatiques. Cela pourrait être fait en calculant, par exemple, les probabilités de transition conditionnelles selon le patron de circulation atmosphérique observé.

BIBLIOGRAPHIE

- Akaike, H. 1974. «A new look at the statistical model identification». In *IEEE Trans. Autom. Control*, vol. 19, p. 716-723.
- Anselin, L. 1980. «Estimation methods for spatial autoregressive structures». *Regional science dissertation and monograph series*, n°8, Cornell University, Ithaca, NY.
- Anselin, L., et A. K. Bera. 1998. «Spatial dependence in linear regression models with an introduction to spatial econometrics». In *Handbook of applied economic statistics*, p. 237-289. Dallas (Texas): D. B. Owen, Founding Editor.
- Bailey, N. T. J. 1964. *The elements of stochastic processes*. Coll. « John Wiley ». New York, 39 p.
- Bardossy, A. 1991. «A stochastic weather generator using atmospheric circulation patterns and its use to evaluate climate change effects». In *proceedings of the international workshop on new uncertainty concepts in hydrology and water resources*, Madralin, Poland.
- Bardossy, A. 1995a. «A stochastic weather generator using atmospheric circulation patterns and its use to evaluate climate change effects». In *new uncertainty concepts in hydrology and water resources*, Ed.: Kundzewicz, C. W., Cambridge University Press, p. 25-31.
- Bardossy, A. 1995b. «Stochastic downscaling of GCM-output results using atmospheric circulation patterns». In *space and time scale variability and interdependencies in hydrological processes, International Hydrology series*, Cambridge University Press, p. 119-123.
- Bardossy, A., et E. J. Plate. 1991. «Modeling daily rainfall using a semi-Markov representation of circulation pattern occurrence». *Journal of Hydrology*, vol. 122, p. 33-47.
- Bardossy, A., et E. J. Plate. 1992. «Space-time models for daily rainfall using atmospheric circulation patterns». *Water resources research*, vol. 28, p. 1247-1259.
- Bellone, E., J. P. Hughes et P. Guttorp. 2000. «A hidden Markov model for downscaling synoptic atmospheric patterns to precipitation amounts». *Climate Research*, vol. 15, p. 1-12.
- Bisson, J.L., et F. Roberge. 1983. *Prévision des apports naturels : expérience d'Hydro-Québec*. Compte-rendu de l'Atelier sur la prévision du débit, Toronto.

- Bogardi, I., I. Matyasovszky, A. Bardossy et L. Duckstein. 1993. «Application of a space-time stochastic model for daily precipitation using atmospheric circulation patterns». *J. Geophys. Res.*, vol. 98, p. 16 653-16 667.
- Brandsma, T., et T. A. Buishand. 1997. «Rainfall generator for the Rhine basin; single-site generation of weather variables by nearest-neighbour resampling». *KNMI-publication 186-1*, 47 p.
- Bras, R., et I. Rodriguez-Iturbe. 1976. «Rainfall generation: a nonstationary time varying multidimensional model». *Water resources research*, vol. 12, p. 450-456.
- Brissette, F., M. Khalili et R. Leconte. 2007. «Efficient stochastic generation of multi-site synthetic precipitation data». *Journal of Hydrology*, in press.
- Bruhn, J. A., W. E. Fry et G. W. Fick, 1980. «Simulation of daily weather data using theoretical probability distributions». *J. Appl. Meteorol.*, vol. 19, p. 1029-1036.
- Buishand, T. A. 1977. «Stochastic modelling of daily rainfall sequences». Dissertation, Wageningen, Mededlingen landbouwhogeschool, vol. 77, n°3, 212 p.
- Buishand, T. A. 1978. «Some remarks on the use of daily rainfall models». *Journal of Hydrology*, vol. 36, p. 295-308.
- Buishand, T. A., et T. Brandsma. 2001. «Multisite simulation of daily precipitation and temperature in the Rhine basin by nearest-neighbour resampling». *Water resources research*, vol. 37, n°11, p. 2761-2776.
- Burger, K. 1958. *Zur Klimatologie der Grosswetterlagen*, Ber. Dtsch. Wetterdienstes 45, vol. 6, Selbstverlag des Deutschen Wetterdienstes, Offenbach am Main, Germany.
- Caron, A. 2006. «Étalonnage et validation d'un générateur de climat dans le contexte des changements climatiques». Mémoire de maîtrise en génie de la construction, Montréal, École de technologie supérieure, 134 p.
- Charles, S. P., B. C. Bates et J. P. Hughes. 1999. «A spatiotemporal model for downscaling precipitation occurrence and amounts». *J. Geophys. Res.*, vol. 104, p. 31657-31669.
- Chin, E. H. 1977. «Modelling daily precipitation occurrence process with Markov chain». *Water Resources Research*, vol. 13, p. 949-956.
- Cliff, A. D., et J. K. Ord. 1981. *Spatial processes: Models and applications*. London: Pion.
- Coe, R., et R. D. Stern. 1982. «Fitting models to rainfall data». *J. appl. Meteorol.*, vol. 21, p. 1024-1031.

- Conway, D., R. L. Wilby et P. D. Jones. 1996. Precipitation and air flow indices over the British Isles. *Clim. Res.*, vol. 7, p. 169-183.
- Corte-Real, J., B. Qian et H. Xu. 1998. «Regional climate change in Portugal : precipitation variability associated with large-scale atmospheric circulation». *International journal of climatology*, vol. 18, p. 619-635.
- Corte-Real, J., B. Qian et H. Xu. 1999b. «Circulation patterns, daily precipitation in Portugal and implications for climate change simulated by the second Hadley Centre GCM». *Climate Dynamics*, vol. 15, p. 921-935.
- Corte-Real, J., H. Xu et B. Qian. 1999a. «A weather generator for obtaining daily precipitation scenarios based on circulation patterns». *Climate Research.*, vol. 13, p. 61-75.
- Cressie, N. 1991. *Statistical analysis of spatial data*, John Wiley and Sons ed., 900 p.
- Cressie, N. A. C. 1993. *Statistics for spatial data*. Wiley series in probability and mathematical statistics, John Wiley & Sons, 900 p.
- Delaunay, D. 1934. Sur la sphère vide, *Izvestia Akademii Nauk SSSR, Otdelenie Matematicheskikh i Estestvennykh Nauk*, vol. 7, p. 793-800.
- Dennett, M.D.J., J.A. Rodgers et J.D.H. Keatinge. 1983. «Simulation of a rainfall record for a new site of a new agricultural development: an example from northern Syria». *Agricultural Meteorology*, vol. 29, p. 247-258.
- Dubrovsky, M., J. Buchtele et Z. Zalud. 2004. «High-frequency and low-frequency variability in stochastic daily weather generator and its effect on agricultural and hydrologic modelling». *Climatic Change*, vol. 63, p. 145-179.
- Efron, B. 1979. «Bootstrap methods: Another look at the Jackknife». *The Annals of Statistics*, vol. 7, n°1, p. 1-26.
- Efron, B., et R. Tibishirani, 1993. *An introduction to the bootstrap*, Chapman & Hall, software. New York.
- Fortin, J. P., R. Turcotte, S. Massicotte, R. Moussa, J. Fitzback et J. P. Villeneuve. 2001a. «Distributed Watershed Model Compatible with Remote Sensing and GIS Data. I: Description of Model». *J. Hydrologic Engrg.*, vol. 6, n°2, p. 91-99.
- Fortin, J. P., R. Turcotte, S. Massicotte, R. Moussa, J. Fitzback et J. P. Villeneuve. 2001b. «Distributed Watershed Model Compatible with Remote Sensing and GIS Data. II: Application to Chaudière Watershed». *J. Hydrologic Engrg.*, vol. 6, n°2, p. 100-108.

- Foufoula-Georgiou, E., et D. P. Lettenmaier. 1987. «A Markov renewal model for rainfall occurrences». *Water resources research*, vol. 23, p. 875-884.
- Gabriel, R. et J. Neuman. 1962. «A Markov chain model for daily rainfall occurrence at Tel Aviv, Israel». *Q. J. R. Meteorol. Soc.*, vol. 88, p. 90-95.
- Gates, P., et H. Tong. 1976. «On Markov chain modelling to some weather data». *J. Appl. Meteorol.*, vol. 15, p. 1145-1151.
- Green, J. R. 1964. «A model for rainfall occurrence». *J. Roy. Statist. Soc.*, vol. B26, p. 345-353.
- Griffith, D. A. 2003. *Spatial autocorrelation and spatial filtering: Gaining understanding through theory and scientific visualization*. Advances in spatial science, Springer, 247 p.
- Gutry-Korycka, M., P. Werner et B. Jakubiak. 1994. «Generation of Time Series of the Meteorological Values in Changing Climatic Conditions». *Geographia Polonica*, vol. 62, p. 23-45.
- Haan, C. T., D. M. Allen et J. O. Street. 1976. «A Markov chain model of daily rainfall». *Water Resources Research*, vol. 12, n°3, p. 443-449.
- Hanson, C.L., K.A. Cumming, D.A. Woolhiser et C.W. Richardson. 1994. *Microcomputer program for daily weather simulation in the contiguous United States* ARS-114. Washington (DC.) : US Department of Agriculture, Agricultural Research Service, 38 p.
- Harmel, R. D., G. Johnson, C. W. Richardson. 2002. «The GEM experience: Weather generator technology development in the USDA». *Bulletin of the American Meteorological Society*, vol. 83, p. 954-957.
- Harmel, R.D., C.W. Richardson et K. W. King. 2000. «Hydrologic response of a small watershed model to generated precipitation». In *Transactions of the ASAE*, vol. 43, p. 1483-1488.
- Harrison, M., et P. Waylen. 2000. «A note concerning the proper choice for Markov model order for daily precipitation in the humid tropics: a case study in Costa Rica». *International Journal of Climatology*, vol. 20, p. 1861-1872.
- Hay, L., G. J. McCabe, D. M. Wolock et M. A. Ayers. 1991. «Simulation of precipitation by weather type analysis». *Water resources research*, vol. 27, p. 493-501.
- Hayhoe, H. N. 1998. «Relationship between weather variables in observed and WXGEN generated data series». *Agricultural and Forest Meteorology*, vol. 90, p. 203-214.

- Hubert, L. J., R. G. Golledge et C. M. Costanzo. 1981. «Generalized procedures for evaluating spatial autocorrelation». *Geographical analysis*, vol. 13, p. 224-233.
- Hughes, J. P., et P. Guttorp. 1994a. «A class of stochastic models for relating synoptic atmospheric patterns to regional hydrologic phenomena». *Water resources research*, vol. 30, p. 1535-1546.
- Hughes, J. P., et P. Guttorp. 1994b. «Incorporating spatial dependence and atmospheric data in a model of precipitation». *J. Appl. Meteorol.*, vol. 33, p. 1503-1515.
- Hughes, J. P., P. Guttorp et S. Charles. 1999. «A nonhomogeneous hidden Markov model for precipitation occurrence». *Appl. Stat.*, vol. 48, p. 15-30.
- Hughes, J. P., D. P. Lettenmaier et P. Guttorp. 1993. «A stochastic approach for assessing the effects of changes in regional circulation patterns on local precipitation». *Water resources research*, vol. 29, p. 3303-3315.
- Jones, W., R. C. Rex et D. E. Threadgill. 1972. «A simulated environmental model of temperature, evaporation, rainfall, and soil moisture». *Trans. ASAE*, vol. 15, p. 366-372.
- Joseph, E.S. 1973. «Spectral analysis of daily maximum and minimum temperature series on the east slope of the Colorado Front Range». *Mon. Weather Rev.*, vol. 101, n°6, p. 505-509.
- Katz, R. W. 1977. «Precipitation as a chain-dependent process». *J. Appl. Meteorol.*, vol. 16, p. 671-676.
- Katz, R.W. 1996. «Use of conditional stochastic models to generate climate change scenarios». *Climatic Change*, vol. 32, p. 237-55.
- Khalili, M., R. Leconte et F. Brissette. 2004a. «Multisite generation of a daily stochastic precipitation to evaluate the effects of the climatic changes on the Châteauguay River Basin hydrology». In *American Geophysical Union, the 2004 Joint Assembly*. (Montréal, mai 17-21 2004).
- Khalili, M., R. Leconte et F. Brissette. 2004b. «Génération multi-site des données de précipitations du bassin versant de la rivière Péribonca à des fins de génération des scénarios de changements climatiques». In *57^e Congrès de l'Association Canadienne des Ressources Hydriques*. (Montréal, juin 16-18 2004).
- Khalili, M., R. Leconte et F. Brissette. 2006. «Efficient watershed modeling using a multi-site weather generator for meteorological data». In *Third International Conference on river basin Management including all aspects of Hydrology, Ecology, Environmental Management, Flood Plains and Wetlands*. (Bologne, Septembre 6-8 2005), p. 273-281. Geo-Environment and Landscape Evolution II, Wessex Institute of Technology, UK, D.

- Khalili, M., R. Leconte et F. Brissette. 2007. «Stochastic multi-site generation of daily precipitation data using spatial autocorrelation». *Journal of hydrometeorology*, vol. 8, issue 3, p. 396-412.
- Khalili, M., F. Brissette et R. Leconte. 2008a. «Stochastic multi-site generation of daily weather data». *Journal of hydrometeorology*, sous évaluation.
- Khalili, M., F. Brissette et R. Leconte. 2008b. «Effectiveness of multi-site weather generator for hydrological modelling: comparative study with uni-site weather generator using distributed hydrological model». *Journal of Hydrology*, sous évaluation.
- Lall, U., et A. Sharma, 1996. «A nearest neighbor bootstrap for resampling hydrologic time series». *Water resources research*, vol. 32, p. 679-693.
- Lana, X., et A. Burgueno. 1998. «Daily dry-wet behaviour in Catalonia (NE Spain) from the viewpoint of Markov chains». *Int. J. Clim.*, vol. 18, p. 793-815.
- Lane, L. J., et M. A. Nearing. 1989. «USDA Water Erosion Prediction Project: Hillslope profile model documentation». *NSERL Rep. 2*, Natl. Soil Erosion Res. Lab., U.S. Dep. Of Agric., Agric. Res. Serv., West Lafayette, Indiana.
- Le Gallo, J. 2000. Econométrie spatiale: autocorrelation spatiale. «Article n°2000-05». France: Université de Bourgogne, 45 p.
- Longley, R. W. 1953. « The length of dry and wet periods ». *Quarterly journal of the royal meteorological society*, vol. 79, p. 520-527.
- Maidment, D. R. 1993. *Handbook of hydrology*. New York : McGraw-Hill.
- Matalas, N. C. 1967. «Mathematical assessment of synthetic hydrology». *Water resources research*, vol. 3, n°4, p. 937-945.
- Mehrotra, R., R. Srikanthan et A. Sharma. 2005. «Comparison of three approaches for stochastic simulation of multi-site precipitation occurrence». *20th Hydrology and Water Resources Symposium*. Canberra.
- Moran, P. A. P. 1950. «Notes on continuous stochastic phenomena». *Biometrika*, vol. 37, p. 17-23.
- Murdoch, J. C., M. Rahmatian et M. A. Thayer. 1993. «A spatially autoregressive median voter model of recreation expenditures». *Public finance quarterly*, vol. 21, p. 334-350.
- Nicks, A. D. 1975. «Stochastic generation of hydrologic model inputs». Ph.D dissertation, Norman, Univ. of Okla. 142 p.

- Nicks, A. D., et J. F. Harp. 1980. «Stochastic generation of temperature and solar radiation data». *J. Hydrol.*, vol. 48, p. 1-7.
- Odland, J. 1988. *Spatial autocorrelation*. Newbury Park: Sage Publications, 87 p.
- Parlange, M. B., et R. W. Katz. 2000. «An extended version of the Richardson model for simulating daily weather variables». *Journal of Applied Meteorology*, vol. 39, p. 610-622.
- Qian, B., J. Corte-Real et H. Xu. 2002. «Multisite stochastic weather models for impact studies». *International journal of climatology*, vol. 22, p. 1377-1397.
- Racsko, P., L. Szeidl et M.A. Semenov. 1991. « A serial approach to local stochastic weather models ». *Ecological Modelling*, vol. 57, p. 27-41.
- Rajagopalan, B., et U. Lall. 1995. «A nearest neighbor bootstrap for resampling daily precipitation and other weather variables», Working paper WP-95-HWR-UL/013.
- Rajagopalan, B., et U. Lall. 1999. «A k-nearest-neighbor simulator for daily precipitation and other weather variables». *Water resources research*, vol. 35, n°10, p. 3089-3101.
- Richardson, C. W. 1981. « Stochastic simulation of daily precipitation, temperature, and solar radiation ». *Water resources research*, vol. 17, n°3, p. 182-190.
- Richardson, C. W., et A. D. Nicks. 1990. «Weather generator description». In Sharpley, A. N. and Williams, J. R., editors, *Erosion/productivity impact calculator. 1. Model documentation. Agricultural research service technical bulletin 1768, Washington, DC: US Department of Agriculture*, p. 93-104.
- Richardson, C. W., et D. A. Wright. 1984. *WGEN: a model of generating daily weather variables*. ARS-8. Washington (DC.): US Department of Agriculture, Agricultural Research Service, 83 p.
- Roldan, J., et D. A. Woolhiser. 1982. «Stochastic daily precipitation models, 1. A comparison of occurrence processes». *Water Resour. Res.*, vol. 18, p. 1451-1459.
- Schwarz, G. 1978. «Estimating the dimension of a model. *Ann. Stat.*, vol. 6, 461-464.
- Selvalingam, S., et M. Miura. 1978. « Stochastic modelling of monthly and daily rainfall sequences ». *Water Resour. Bull.*, vol. 14, p. 1105-1120.
- Semenov, M. A., et E. M. Barrow. 1997. « Use of a stochastic weather generator in the development of climate change scenarios ». *Climatic change*, vol. 22, p. 67-84.
- Semenov, M.A., R.J. Brooks, E.M. Barrow et C.W. Richardson. 1998. «Comparison of WGEN and LARS-WG stochastic weather generators for diverse climates». *Climate Research*, vol. 10, p. 95-107.

- Singh, S. V., R. H. Kripalani, P. Saha, P. M. M. Ismail et S. D. Dahale. 1981. «Persistence in daily and 5-day summer monsoon rainfall over India». *Arch. Met. Geoph. Biokl., Ser. A*, vol. 30, p. 261-277.
- Siriwardena, L., R. Srikanthan et T. A. McMahon. 2002. *Evaluation of two daily rainfall data generation models*. «Technical Report 02/14», CRC for Catchment Hydrology.
- Smith, J. A., et A. F. Karr. 1985. «Parameter estimation for a model of space-time rainfall». *Water Resources Research*, vol. 20, n°8, p. 1251 - 1257.
- Smith, R. E., et H. A. Schreiber. 1973. «Point processes of seasonal thunderstorm rainfall, 1, Distribution of rainfall events». *Water Resources Research*, vol. 9, n°4, p. 871-884.
- Srikanthan, R., et T. A. McMahon. 2001. «Stochastic generation of annual, monthly and daily climate data: A review». *Hydrology and Earth System Sciences*, vol. 5, n°4, p. 653-670.
- Stern, R. D., et R. Coe. 1984. «A model fitting analysis of daily rainfall data». *Journal of the royal statistical society series A*, vol. 147, part 1, p. 1-34.
- Thom, H. C. S. 1958. «A note on the gamma distribution». *Mon. Wea. Rev.*, vol. 86, p. 117-122.
- Tobler, W. 1970. «A computer movie simulating urban growth in the Detroit region». *Economic geography*, vol. 46, p. 234-240.
- Todorovic, P., et D. A. Woolhiser. 1975. «A stochastic model of n-day precipitation». *J. Appl. Meteorol.*, vol. 14, p. 17-24.
- Upton, G. J. G., et B. Fingleton. 1985. *Spatial data analysis by example: point Pattern and Interval Data*. John Wiley, 410 p.
- Vogel, R. M., et A. L. Shallcross. 1996. «The moving blocks bootstrap versus parametric time series models». *Water Resources Research*, vol. 32, p. 1875-1882.
- Wallis, T. W. R., et J. F. Griffiths. 1997. «Simulated meteorological input for agricultural models». *Agricultural and Forest Meteorology*, vol. 88, p. 241-258.
- Watson, B.M., R. Srikanthan, S. Selvalingam et M. Ghafouri. 2005. «Hydrologic response of SWAT to single site and multi-site daily rainfall generation models». In *Proceedings of MODSIM05 International Congress on Modelling and Simulation*. Melbourne, Australia, p. 2981-2987.

- Waymire, E., V.K. Gupta et I. Rodriguez-Iturbe. 1984. «A spectral theory of rainfall intensity at the meso- β scale». *Water Resources Research*, vol. 20, p. 1453-1465.
- Wilby, R. L. 1994. «Stochastic weather type simulation for regional climate change impact assessment». *Water resources research*, vol. 30, p. 3395-3403.
- Wilks, D. S. 1989. «Conditioning stochastic daily precipitation models on total monthly precipitation». *Water resources research*, vol. 25, p. 1429-1439.
- Wilks, D. S. 1992. «Adapting stochastic weather generation algorithms for climate change studies». *Climatic change*, vol. 22, p. 67-84.
- Wilks, D. S. 1995. *Statistical methods in the atmospheric sciences*. International Geophysics Series, vol. 59, Academic Press, 464 p.
- Wilks, D. S. 1997. «Resampling hypothesis tests for autocorrelated fields». *J. Climate*, vol. 10, p. 65-82.
- Wilks, D. S. 1998. «Multisite generalization of a daily stochastic precipitation generation model». *Journal of Hydrology*, vol. 210, p. 178-191.
- Wilks, D. S. 1999a. «Multisite downscaling of daily precipitation with a stochastic weather generator». *Climate Research*, vol. 11, p. 125-136.
- Wilks, D. S. 1999b. «Simultaneous stochastic simulation of daily precipitation, temperature and solar radiation at multiple sites in complex terrain». *Agricultural and Forest Meteorology*, vol. 96, p. 85-101.
- Wilks, D. S., et R. L. Wilby. 1999. «The weather generation game: a review of stochastic weather models». *Progress in physical geography*, vol. 23, n°3, p. 329-357.
- Williams, C. B. 1952. «Sequences of wet and of dry days considered in relation to the logarithmic series ». *Quarterly Journal of the Royal Meteorological Society*, vol. 78, p. 91-96.
- Woolhiser, D. A., et G. G. S. Pegram. 1979. «Maximum likelihood estimation of Fourier coefficients to describe seasonal variations of parameters in stochastic daily precipitation models». *J. Appl. Meteorol.*, vol. 18, n°1, p. 34-42.
- Woolhiser, D. A., et J. Roldan. 1982. «Stochastic daily precipitation models, 2: A comparison of distribution of amounts». *Water resources research*, vol. 18, p. 1461-1468.
- Woolhiser, D. A., et J. Roldan. 1986. «Seasonal and regional variability of parameters for stochastic daily precipitation models». *Water resources research*, vol. 22, p. 965-978.

- Yakowitz, S. J. 1979. «A nonparametric Markov model for daily river flow». *Water resources research*, vol. 15, p. 1035-1043.
- Young, K. C. 1994. «A multivariate chain model for simulating climatic parameters from daily data». *J. Appl. Meteorol.*, vol. 33, p. 661–671.
- Zorita, E., J. P. Hughes, D. P. Lettenmaier et H. V. Storch. 1995. «Stochastic characterization of regional circulation patterns for climate model diagnosis and estimation of local precipitation». *J. Climate*, vol. 8, p. 1023-1042.
- Zucchini, W., et P. Guttorp. 1991. «A hidden Markov model for space-time precipitation». *Water resources research*, vol. 27, p. 1917-1923.